# Maths Finance (2020/21)
# Block I

Ellen Powell
*Please send comments and corrections to ellen.g.powell@durham.ac.uk*
(See references for complementary course texts)

## Contents

# Timetable

Each of the four weeks commencing 5/10, 12/10, 19/10, 26/10 there will be:

- Three hour-long content based activities, that will be uploaded to DUO on **Monday, Tuesday and Wednesday mornings by 10am**. These will consist of videos to watch and related problems to solve.
- Two online office hours on **Tuesdays 09:00-10:00 & 12:00-13:00**, where I will be available on Zoom to answer any questions about the course.
- A problems session on **Fridays 09:00-10:00**, also on Zoom, where I will take requests to go over problems from the set for that week. Solutions will be available afterwards.

**Zoom meeting details**

https://durhamuniversity.zoom.us/j/99149091697?pwd=S2o4N3dNcFFlcW1HOUo3K3duT2ZnQT09
Meeting ID: 991 4909 1697
Passcode: 934060

Outside of these times, do not hesitate to email me at ellen.g.powell@durham.ac.uk with any questions you may have. I will endeavour to respond as soon as possible. If I do not respond within 24 hours, feel free to send me a reminder!

## Week 1

**Content 1:** §1.1
**Content 2:** §1.2
**Content 3:** §1.3
**Problems session:** Problems from §1

## Week 2

**Content 1:** §2.1
**Content 2:** §2.2
**Content 3:** §3.1
**Problems session:** Problems from §2

## Week 3

**Content 1:** §3.2
**Content 2:** §3.3
**Content 3:** §4.1
**Problems session:** Problems from §3

## Week 4

**Content 1:** §4.2
**Content 2:** §5.1
**Content 3:** §5.2
**Problems session:** Problems from §4 and §5

# 1 Probability spaces and random variables
Complementary reading: [1, §1.1] and [1, Appendix 2].

## 1.1 Probability spaces

**Definition 1.1** (Probability space)**.** *A **probability space** or **probability triple** is a collection* $(\Omega, \mathcal{F}, \mathbb{P})$ *where:*

- $\Omega$ *is the **sample space** (the set of all possible **outcomes**);*
- $\mathcal{F}$ *is a $\sigma$-**field** (consisting of **events**);*
- $\mathbb{P}$ *is a **probability measure**, assigning probabilities to events in* $\mathcal{F}$*.*

**Definition 1.2** ($\sigma$-field)**.** *A $\sigma$-field (or $\sigma$ algebra) $\mathcal{F}$ on $\Omega$ is a collection of subsets of $\Omega$, such that*

*(i)* $\emptyset \in \mathcal{F}$,

*(ii) if $A \in \mathcal{F}$, then $A^c \in \mathcal{F}$, and*

*(iii) if $(A_i)_{i \geq 1} \in \mathcal{F}$ then $\cup_{i=1}^{\infty} A_i \in \mathcal{F}$.*

So in words, a $\sigma$-field is a collection of subsets of the sample space, that contains the empty set and is closed under taking complements and unions.

Here are some simple examples of $\sigma$-fields.

- $\{\emptyset, \Omega\}$. This is the smallest possible $\sigma$-field, since by definition a $\sigma$-field has to contain the empty set and its complement
- $\mathcal{P}(\Omega)$, the set of all subsets of $\Omega$. This is clearly the biggest possible $\sigma$-field
- $\{\emptyset, \Omega, A, A^c\}$ for some $A \subset \Omega$. This is the $\sigma$-field "generated by the event $A$".

In general, if $\mathcal{C}$ is a collection of subsets of $\Omega$ (now not necessarily satisfying $(i) - (iii)$ from Definition 1.2), it is possible to define the **$\sigma$-field generated by $\mathcal{C}$.** This is the **smallest** $\sigma$-field that contains all the elements of $\mathcal{C}$, and is denoted by $\sigma(\mathcal{C})$. Equivalently, $\sigma(\mathcal{C})$ is the intersection of all $\sigma$-fields that contain $\mathcal{C}$. Clearly if $\mathcal{C}$ already satisfies the axioms of Definition 1.2, then $\sigma(\mathcal{C}) = \mathcal{C}$.

**Example 1.3.** *Let $\Omega = \mathbb{R}$ and $\mathcal{C} := \{(a, b) : -\infty < a < b < +\infty\}$. Then the $\sigma$-field $\sigma(\mathcal{C})$ generated by $\mathcal{C}$, is called the **Borel $\sigma$-field** on $\mathbb{R}$. This will be denoted by $\mathfrak{B}(\mathbb{R})$, and sets in $\mathfrak{B}(\mathbb{R})$ will be called **Borel sets**.*

**Remark 1.4.** In general, if $\mathcal{T} \subset \mathbb{R}$ then the **Borel $\sigma$-field** on $\mathcal{T}$, denoted $\mathfrak{B}(\mathcal{T})$, is the $\sigma$-field generated by open sets with respect to $\mathcal{T}$.

**Definition 1.5** (Probability measure)**.** *Let $(\Omega, \mathcal{F})$ be a set and a $\sigma$-field on it. A **probability measure** $\mathbb{P}$ is a function*

$$\mathbb{P} : \mathcal{F} \to [0, 1], \tag{1}$$

*satisfying the **probability axioms**. Equivalently, a number $\mathbb{P}(A)$ (the "probability of $A$") assigned to each $A \in \mathcal{F}$. The probability axioms are:*

*(i) $\mathbb{P}(\Omega) = 1$ (the set of all outcomes has probability one);*

*(ii) if $(A_i)_{i \geq 1} \in \mathcal{F}$ and $A_i \cap A_j = \emptyset$ for $i \neq j$, then*

$$\mathbb{P}(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mathbb{P}(A_i) \tag{2}$$

*(the probability of a disjoint union of events is the sum of the individual probabilities).*

If an event $A \subset \mathcal{F}$ satisfies $\mathbb{P}(A) = 1$, then it is said that $A$ occurs **almost surely** (or a.s. for short).

**Example 1.6.** *Suppose you want to set up a probability space for a fair die roll. Then it is natural to define the sample space, the set of possible outcomes, as $\Omega = \{1, 2, 3, 4, 5, 6\}$ and to set $\mathcal{F} = \mathcal{P}(\Omega)$. Since the die is fair, the probablity measure $\mathbb{P}$ should be defined so that $\mathbb{P}(\{i\}) = 1/6$ for $i = 1, 2, 3, 4, 5, 6$. Note that this defines $\mathbb{P}(A)$ for any $A \in \mathcal{F}$ by axiom (ii) of Definition 1.5, since any $A \in \mathcal{F}$ is equal to $\cup_{i\in I}\{i\}$ for some subset $I$ of $\{1, 2, 3, 4, 5, 6\}$.*

**Example 1.7.** *Let $\Omega = [0, 1]$ and $\mathfrak{B}([0, 1])$ be the Borel $\sigma$-field. Define $\mathbb{P}$ by setting*

$$\mathbb{P}((a, b)) = b - a \tag{3}$$

*for any open set $(a, b) \subset [0, 1]$. (This uniquely defines $\mathbb{P}(A)$ for any $A \in \mathcal{B}([0, 1])$ by **Carathéodory's extension theorem**, but we will not go into this here). Then $\mathbb{P}$ is a probability measure (called **Lebesgue measure** or **uniform measure** on $[0, 1]$) and $([0, 1], \mathfrak{B}([0, 1]), \mathbb{P})$ is a probability space.*

One important notion in probability theory is that of **independence**. Roughly speaking, events are independent if the occurrence of one of them does not affect the likelihood of the other occurring.

**Definition 1.8** (Independence of events)**.** *Suppose $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability triple. Then $A, B \subset \mathcal{F}$ are said to be **independent** if*

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B). \tag{4}$$

**Example 1.9.** *Suppose a fair coin is tossed twice. Then the events $A_1 =$\{coin 1 is heads\} and $A_2 =$\{coin 2 is heads\} are independent, and $\mathbb{P}(A_1 \cap A_2) = 1/2 \times 1/2 = 1/4$. But if $A_3 =$\{both coins are tails\} then $A_1$ and $A_3$ (similarly $A_2$ and $A_3$) are not independent, because clearly $\mathbb{P}(A_3 \cap A_1) = 0 \neq \mathbb{P}(A_3)\mathbb{P}(A_1)$.*

More generally, one can describe how the occurrence of one event affects the likelihood of another occurring. This gives rise to the definition of **conditional probability**.

**Definition 1.10** (Conditional probability)**.** *Suppose $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability triple and $A, B \subset \mathcal{F}$. The **conditional probability of** $A$ **given** $B$ is defined as*

$$\mathbb{P}(A|B) := \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}. \tag{5}$$

**Example 1.11.** *Returning to Example 1.9, this definition gives that $\mathbb{P}(A_3|A_1) = 0/\mathbb{P}(A_1) = 0$, which makes sense, since intuitively "the probability that both coins are tails given the first is heads" really should be $0$. A slightly more complicated example would be to work out $\mathbb{P}(A_3|A_1^c)$. That is, the probability that both coins are tails given the first coin is tails. Intuitively this should be $1/2$ - the probability that the second coin is tails - since if that happens then clearly both coins will be tails. Checking with the formula:*

$$\mathbb{P}(A_3|A_1^c) = \frac{\mathbb{P}(A_3 \cap A_1^c)}{\mathbb{P}(A_1^c)} = \frac{\mathbb{P}(A_1^c \cap A_2^c)}{\mathbb{P}(A_1^c)} = \frac{1/2 \times 1/2}{1/2} = \frac{1}{2} \tag{6}$$

*as expected.*

### Exercises for §1.1

**Exercise 1.1.** *Let $\mathcal{F}_1$ and $\mathcal{F}_2$ be $\sigma$-fields on $\Omega$. Show that $\mathcal{F}_1 \cap \mathcal{F}_2$ is a $\sigma$-field. If $\{\mathcal{F}_\alpha; \alpha \in \mathcal{A}\}$ are a collection of $\sigma$-fields indexed by some set $\mathcal{A}$, show that $\cap_{\alpha\in\mathcal{A}}\mathcal{F}_\alpha$ is always a $\sigma$-field.*

**Exercise 1.2.** *Let $\Omega = \{\omega_1, \omega_2, \omega_3\}$. By considering suitable $\sigma$-fields on $\Omega$, show that the union of two $\sigma$-fields is not necessarily a $\sigma$ field.*

**Exercise 1.3.** *Describe $\sigma(\mathcal{C})$ when $\mathcal{C}$ is a finite collection $A_1, \ldots, A_n \subset \Omega$.*

**Exercise 1.4.** *Show that $\mathfrak{B}(\mathbb{R})$ is also equal to $\sigma(\tilde{\mathcal{C}})$ where $\tilde{\mathcal{C}} = \{(-\infty, c]; c \in \mathbb{R}\}$*

**Exercise 1.5.** *If $A$ and $B$ are any two events, show that $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$. If $A, B, C$ are any three events, show that*

$$\mathbb{P}(A^c \cap (B \cup C)) = \mathbb{P}(B) + \mathbb{P}(C) - \mathbb{P}(A \cap B) - \mathbb{P}(B \cap C) - \mathbb{P}(C \cap A) + \mathbb{P}(A \cap B \cap C). \tag{7}$$

*Use this result to deduce how many of the numbers from 1 to 500 are not divisible 7, but are divisible by 3 or 5.*

**Exercise 1.6.** *Three machines, A, B and C, produce components. 10% of components from A are faulty, 20% of components from B are faulty, and 30% of components from C are faulty. Equal numbers from each machine are collected in a packet. (i) One component is selected at random from the packet. What is the probability that it is faulty? (ii) Suppose a component is drawn from the packet and found to be faulty. What is the probability that it was made by machine A?*

### 1.2 Random variables

**Definition 1.12** (Measurable functions). *Let $\mathcal{F}$ be a $\sigma$-field on $\Omega$. A function $X : \Omega \to \mathbb{R}$ is said to be $\mathcal{F}$-**measurable** if*

$$\{\omega \in \Omega : X(\omega) \in A\} \in \mathcal{F} \text{ for all } A \in \mathfrak{B}(\mathbb{R}), \tag{8}$$

*equivalently*

$$\{\omega \in \Omega : X(\omega) \leq x\} \in \mathcal{F} \text{ for all } x \in \mathbb{R}. \tag{9}$$

The two definitions above are equivalent by Exercise 1.4.

**Definition 1.13** (Random Variables). *If $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space, then a **random variable** on $(\Omega, \mathcal{F}, \mathbb{P})$ is just an $\mathcal{F}$-measurable function*

$$X : \Omega \to \mathbb{R}. \tag{10}$$

**Example 1.14.** *An important example of a random variable is the **indicator function** of an event $A \in \mathcal{F}$. This is the $\mathcal{F}$-measurable function $\mathbf{1}_A : \Omega \to \mathbb{R}$ defined by*

$$\mathbf{1}_A(\omega) = \begin{cases} 1 & \omega \in A \\ 0 & \omega \notin A \end{cases} \tag{11}$$

**Example 1.15.** *Consider as in Example 1.6 the probability space for a fair die roll. Then $X(\omega) = \omega$ (the score) and $Y(\omega) = \mathbf{1}_{\{\omega \text{ is even}\}}$ (the indicator that the score is even) are random variables on $(\Omega, \mathcal{F}, \mathbb{P})$, since they are clearly $\mathcal{F}$-measurable.*

The random variables in this course will come in two flavours: **discrete** and **continuous**. A **discrete** random variable is one that can only take values in some countable set $\mathcal{X}$. Such a random variable can be described completely by its **probability mass function**: that is, the countable collection of numbers

$$\{p_x\,;\, x \in \mathcal{X}\} \text{ with } \sum_{x \in \mathcal{X}} p_x = 1, \tag{12}$$

where $p_x = \mathbb{P}(\{X = x\})$ is the probability of the event that $X$ is equal to $x$. A continuous random variable is one that can take more than a countable collection of possible values. This is described by its **cumulative distribution function**

$$F : \mathbb{R} \to [0, 1]\,;\; F(x) = \mathbb{P}(\{X \leq x\}) \text{ for } y \in \mathbb{R}. \tag{13}$$

If there is exists a function $p$ such that

$$F(x) = \mathbb{P}(\{X \le x\}) = \int_{-\infty}^{x} p(y)\, dy \tag{14}$$

for every $x$, then $p$ is called the **probability density function** of $X$. This is the "infinitesimal" analogue of the probability mass function: when $dx$ is very small $\mathbb{P}(X \in [x, x+dx]) \approx p(x)dx$.

The **distribution** or **law** of a random variable $X$ is just the information $\mathbb{P}(X \in A)$ for all $A \in \mathfrak{B}(\mathbb{R})$.

**Example 1.16.** *The random variables $X, Y$ from Example 1.15 are both discrete random variables since they can only take values in $\{1, \dots, 6\}$ and $\{0, 1\}$. The probability density function of $X$, for instance, is given by $p_i = 1/6$ for $i = 1, \dots, 6$.*

*On the other hand, consider the probability space $([0,1], \mathfrak{B}([0,1]), \mathbb{P})$ from Example 1.7. Then the function $Z : \Omega \to \mathbb{R}$ defined by $U(\omega) = \omega$ for $\omega \in [0,1]$ is a continuous random variable called the* ***uniform*** *random variable on $[0,1]$.*

**Definition 1.17** (Measurability)**.** *Suppose that $X$ is a random variable on $(\Omega, \mathcal{F})$ and $\mathcal{G} \subset \mathcal{F}$ is a $\sigma$-field that is contained in the $\sigma$-field $\mathcal{F}$. Then $X$ is said to be $\mathcal{G}$-measurable if*

$$\{\omega \in \Omega : X(\omega) \le x\} \in \mathcal{G} \text{ for all } x \in \mathbb{R}. \tag{15}$$

*Roughly speaking, $X$ is $\mathcal{G}$-measurable if "knowing $\mathcal{G}$" means "knowing $X$".*

**Example 1.18.** *Take $\mathcal{G} \subset \mathcal{F}$ given by*

$$\mathcal{G} = \{\emptyset, E, E^{c}, \Omega\}, \tag{16}$$

*where $E = \{2, 4, 6\}$ is the event that the score is even, and its complement $E^{c} = \{1, 3, 5\}$ (the score is odd). Note that $\mathcal{G}$ is a $\sigma$-algebra.*

*Moreover, $Y$ is $\mathcal{G}$-measurable, since e.g. $\{\omega : Y(\omega) \le 1/2\} = E^{c} \in \mathcal{G}$, but $X$ is* **not** *$\mathcal{G}$-measurable, since e.g. $\{\omega : X(\omega) \le 1\} = \{1\} \notin \mathcal{G}$.*

Just as with subsets of outcomes, it is possible to define the $\sigma$-field generated by a random variable $X$. Informally this is the $\sigma$-field "containing all the information about the random variable $X$".

**Definition 1.19** ($\sigma$-field generated by a random variable)**.** *The* ***$\sigma$-field generated by a random variable*** *$X : \Omega \to \mathbb{R}$ consists of all sets of the form $\{\omega : X(\omega) \in A\}$ for $A \in \mathfrak{B}(\mathbb{R})$. This $\sigma$-field will be denoted by $\sigma(X)$.*

**Definition 1.20** (Independence of $\sigma$-fields and random variables)**.** *Two $\sigma$-fields $\mathcal{G}_1 \subset \mathcal{F}$ and $\mathcal{G}_2 \subset \mathcal{F}$ are said to be independent if any two events $A_1 \in \mathcal{G}_1$ and $A_2 \in \mathcal{G}_2$ are independent. A random variable $X$ is said to be independent of a $\sigma$-field $\mathcal{G} \subset \mathcal{F}$ if $\sigma(X)$ and $\mathcal{G}$ are independent. Similarly, two random variables $X, Y$ are said to be independent if $\sigma(X)$ and $\sigma(Y)$ are independent.*

This definition can be extended to families of $\sigma$-algebras and random variables. For example, if $\mathcal{G}_1, \mathcal{G}_2, \dots$ are a family of sub $\sigma$-fields on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ then they are said to be independent if

$$\mathbb{P}(A_1 \cap \dots \cap A_n) = \mathbb{P}(A_1) \cdots \mathbb{P}(A_n) \tag{17}$$

whenever $k_1, \dots, k_n \ge 1$ and $A_i \in \mathcal{G}_{k_i}$ for each $1 \le i \le n$.

**Example 1.21.** *If $X$ and $Y$ are two discrete random variables, taking values in $\mathcal{X}$ and $\mathcal{Y}$ respectively, then they are independent if and only if*

$$\mathbb{P}(X = x, Y = y) = \mathbb{P}(X = x)\mathbb{P}(Y = y) \tag{18}$$

*for all $x \in \mathcal{X}$, $y \in \mathcal{Y}$.*

Let us conclude this subsection with a couple more straightforward definitions.

**Definition 1.22** (Random vector)**.** *A* ***random vector*** *on* $(\Omega, \mathcal{F})$ *is just a vector* $(X_1, \ldots, X_n)$ *where each* $X_i$ *is a random variable.*

**Definition 1.23** (Independent and identically distributed random variables)**.** *A sequence* $(X_i)_{i\geq 1}$ *of random variables on a probability space* $(\Omega, \mathcal{F}, \mathbb{P})$ *are said to be* ***independent and identically distributed*** *or* ***i.i.d.*** *if they are independent and each have the same distribution. That is, for any* $A \in \mathfrak{B}(\mathbb{R})$, $\mathbb{P}(X_i \in A)$ *does not depend on* $i$.

For example, the outcomes of successive fair die rolls would form an i.i.d. sequence.

### Exercises for §1.2

**Exercise 1.7.** *Determine the distribution function and probability density function of* $U$ *from Example 1.7.*

**Exercise 1.8.** *Describe the* $\sigma$*-field generated by* $Y$ *from example Example 1.15.*

**Exercise 1.9.** *Check the conclusion of Example 1.21. Are* $X$ *and* $Y$ *from Example 1.15 independent?*

## 1.3 Moments

### Expectation

**Definition 1.24** (Expectation: positive, discrete case)**.** *Suppose that* $X$ *is a discrete random variable taking values in a countable subset* $\mathcal{X}$ *of* $\mathbb{R}$*, and that* $X$ *is* ***non-negative****, i.e. all elements of* $\mathcal{X}$ *are* $\geq 0$*. Write* $(p_x; x \in \mathcal{X})$ *for the probability mass function of* $x$*. Then the expectation of* $X$ *is defined by*

$$\mathbb{E}(X) := \sum_{x\in\mathcal{X}} x.p_x \in [0, \infty] \tag{19}$$

*Note that if* $\mathcal{X}$ *is infinite, the sum can be defined as the limit as* $n \to \infty$ *of* $a_n := \sum_{x\in\mathcal{X}_n} x.p_x$*, where* $\mathcal{X}_n$ *is increasing in* $n$ *and finite for each* $n$ *with* $\cup_n \mathcal{X}_n = \mathcal{X}$*. The sequence* $a_n$ *is increasing in* $n$ *(since all the* $x.p_x$ *are non-negative) and so has a limit as* $n \to \infty$*. However, this* ***may be infinite****.*

**Definition 1.25** (Expectation: positive, general case)**.** *In general if* $(\Omega, \mathcal{F}, \mathbb{P})$ *is a probability space and* $X$ *is a non-negative random variable on* $\Omega$ *- that is,* $\mathbb{P}(X < 0) = 0$ *- define*

$$\mathbb{E}(X) = \lim_{n\to\infty} \mathbb{E}(2^{-n}\lfloor 2^n X\rfloor) \in [0, \infty] \tag{20}$$

*which exists since the sequence* $\lim_{n\to\infty} \mathbb{E}(2^{-n}\lfloor 2^n X\rfloor)$ *is increasing. Note that* $2^{-n}\lfloor 2^n X\rfloor$ *is a non-negative discrete random variable for every* $X$ *so each expectation in the sequence makes sense by the previous definition. Again this limit* ***may be infinite****.*

**Definition 1.26** (Expectation: general case)**.** *If* $X$ *is not assumed to be non-negative, it is possible to write*

$$X = X^+ - X^- := X\mathbf{1}_{X\geq 0} + X\mathbf{1}_{X<0} \tag{21}$$

*where* $X^\pm$ *are both non-negative.* $X$ *is then said to be* ***integrable*** *if* $\mathbb{E}(X^+)$ *and* $\mathbb{E}(X^-)$ *are both finite. In this case, the expectation of* $X$ *is defined by*

$$\mathbb{E}(X) = \mathbb{E}(X^+) - \mathbb{E}(X^-). \tag{22}$$

**Example 1.27.** *If* $X$ *takes values in a finite set* $\mathcal{X}$*, this is easy.* $X$ *is always integrable and* $\mathbb{E}(X) = \sum_{x\in\mathcal{X}} x.p_x$*. For example, if* $X$ *is the outcome of a fair dice roll as in Example 1.15 then* $\mathbb{E}(X) = (1/6).1 + (1/6).2 + \ldots + (1/6).6 = 21/6$.

The expectation of a random variable $X$ is often referred to as its **mean**.

**Example 1.28.** *If $X$ is a continuous random variable with probability density function $p$ then $X$ is integrable if and only if*

$$\int_0^\infty xp(x)\,dx < \infty \text{ and } \int_{-\infty}^0 xp(x)\,dx < \infty \tag{23}$$

*with*

$$\mathbb{E}(X) = \int_{-\infty}^\infty xp(x)\,dx. \tag{24}$$

*(This can be shown using the definitions but is omitted here).*

**Proposition 1.29** (Properties of expectation)**.** *(i) If $\mathbb{P}(X \geq 0) = 1$ then $\mathbb{E}(X) \geq 0$.*

*(ii) If $X = a$ with $a \in \mathbb{R}$ is a constant random variable then $\mathbb{E}(X) = a$.*

*(iii) If $X$ is a random variable and $c \in \mathbb{R}$ then $\mathbb{E}(cX) = c\mathbb{E}(X)$.*

*(iv) For any finite sequence $X_1, \ldots, X_n$ of random variables*

$$\mathbb{E}(\sum_i X_i) = \sum_i \mathbb{E}(X_i). \tag{25}$$

*(v) If $X$ and $Y$ are independent, integrable random variables, then*

$$\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y). \tag{26}$$

*(vi) For any integrable $X$, $|\mathbb{E}(X)| \leq \mathbb{E}(|X|)$.*

**Lemma 1.30** (Markov's inequality)**.** *If $X$ is a random variable and $a \in \mathbb{R}$, then*

$$\mathbb{P}(|X| \geq a) \leq \frac{\mathbb{E}(|X|)}{a}. \tag{27}$$

*Proof.* Observe that $a\mathbf{1}_{|X|\geq a} \leq |X|$ with probability one. Therefore by Proposition 1.29(i) it follows that $\mathbb{E}(a\mathbf{1}_{|X|\geq a}) \leq \mathbb{E}(|X|)$ and by Proposition 1.29(iii) that $a\mathbb{E}(\mathbf{1}_{|X|\geq a}) \leq \mathbb{E}(|X|)$. Applying the conclusion of Exercise 1.10 from this section, and rearranging, gives the result. □

**Lemma 1.31** (Law of the unconscious statistician (LOTUS))**.** *If $X$ is a discrete random variable with probability mass function $(p_x : x \in \mathcal{X})$ and $f$ is a function from $\mathcal{X} \to [0, \infty)$, then*

$$\mathbb{E}(f(X)) = \sum_{x\in\mathcal{X}} f(x)\,p_x \in [0, \infty]. \tag{28}$$

*(If $f$ can be negative, write $f = f^+ - f^-$ with $f^+, f^-$ non-negative. Then if $\mathbb{E}(f^\pm(X)) < \infty$ the above still holds.)*

*If $X$ is a continuous random variable with probability density function $p : \mathbb{R} \to [0, 1]$ and $f$ is a measurable function from $\mathfrak{B}(\mathbb{R}) \to [0, \infty)$, then $f(X)$ is a random variable and*

$$\mathbb{E}(f(X)) = \int_{\mathbb{R}} f(x)\,p(x)\,dx \in [0, \infty]. \tag{29}$$

*Proof.* Suppose that $X$ is discrete, taking values in the countable set $\mathcal{X} \subset \mathbb{R}$. Then $f(X)$ is discrete taking values in the countable set $f(\mathcal{X}) := \{f(x) : x \in \mathcal{X}\}$. So

$$\mathbb{E}(f(X)) = \sum_{y\in f(\mathcal{X})} y \sum_{\substack{x\in\mathcal{X} \text{ s.t.} \\ f(x)=y}} p_x = \sum_{x\in\mathcal{X}} p_x \big( \sum_{\substack{y\in f(\mathcal{X}) \text{ s.t.} \\ f(x)=y}} y \big) = \sum_{x\in\mathcal{X}} f(x)\,p_x \tag{30}$$

as required. □

**Example 1.32.** *The **exponential distribution** describes the law of a commonly used continuous random variable. If $Y$ is exponentially distributed with parameter $\lambda > 0$, written $Y \sim \text{Exp}(\lambda)$, then $Y$ has probability density function $p(y) = \lambda \exp(-\lambda y)\mathbf{1}_{y\geq 0}$. The expectation of $Y$ can then be calculated as:*

$$\mathbb{E}(Y) = \int_0^\infty \lambda y \exp(-\lambda y)\, dy = [-y\exp(-\lambda y)]_0^\infty + \int_0^\infty \exp(-\lambda y)\, dy = \frac{1}{\lambda}. \tag{31}$$

## Variance

**Definition 1.33** (Variance)**.** *A random variable $X$ is said to have **finite variance**, or be **square integrable**, if*

$$\mathbb{E}(X^2) < \infty. \tag{32}$$

*If this holds then its variance is defined by*

$$\text{var}(X) = \mathbb{E}((X - \mathbb{E}(X))^2) = \mathbb{E}(X^2) - \mathbb{E}(X)^2. \tag{33}$$

**Example 1.34.** *If $X$ is **bounded**, for example if $X$ is the uniform random variable on $[0,1]$ from Example 1.7, then $X^2$ is bounded and so $X$ must be square-integrable.*

More generally, a random variable $X$ is said to have **finite $p$th moment** if $\mathbb{E}(|X|^p) < \infty$.

**Proposition 1.35** (Properties of variance)**.** *(i)* $\text{var}(X) \geq 0$ *with equality if and only if $X$ is constant.*

*(ii)* $\text{var}(aX) = a^2 \text{var}(X)$ *for $a \in \mathbb{R}$.*

*(iii) If $X, Y$ are independent then* $\text{var}(X + Y) = \text{var}(X) + \text{var}(Y)$.

**Lemma 1.36** (Cauchy–Schwarz inequality)**.** *If $X, Y$ are square integrable random variables then*

$$\mathbb{E}(|XY|) \leq \sqrt{\mathbb{E}(X^2)\mathbb{E}(Y^2)}. \tag{34}$$

*Proof.* Observe that for any $a \in \mathbb{R}$, $\mathbb{E}((X - aY)^2) = \mathbb{E}(X^2) - 2a\mathbb{E}(XY) + a^2\mathbb{E}(Y^2) \geq 0$. Setting $a = \mathbb{E}(XY)/\mathbb{E}(Y^2)$ gives that $\mathbb{E}(X^2) - \mathbb{E}(XY)/\mathbb{E}(Y^2) \geq 0$ which provides the result. □

## Covariance and correlation

Covariance and correlation are ways of describing "how dependent" different random variables are.

**Definition 1.37** (Covariance)**.** *Suppose that $X$ and $Y$ are two square-integrable random variables. Then the **covariance** between them is defined by*

$$\text{cov}(X, Y) = \mathbb{E}((X - \mathbb{E}(X))(X - \mathbb{E}(Y)) \tag{35}$$

Note that for **independent random variables**

$$\text{cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y) = \mathbb{E}(X)\mathbb{E}(Y) - \mathbb{E}(X)\mathbb{E}(Y) = 0. \tag{36}$$

So roughly speaking, a bigger covariance means "more dependence". But this isn't perfect, since a small covariance could just be because the random variables themselves do not vary very much (even if they depend heavily on each other). To get around this, the following definition is also often used.

**Definition 1.38** (Correlation)**.** *If $X, Y$ are square integrable random variables with* $\text{var}(X) > 0, \text{var}(Y) > 0$*, then the **correlation** between $X$ and $Y$ is defined by*

$$\text{corr}(X, Y) = \frac{\text{cov}(X, Y)}{\sqrt{\text{var}(X)}\sqrt{\text{var}(Y)}}. \tag{37}$$

By the Cauchy–Schwarz inequality, Lemma 1.36, it holds that $\text{corr}(X, Y) \in [-1, 1]$.

### Exercises from §1.3

**Exercise 1.10.** *Show that if* $Y = \mathbf{1}_A$ *is the indicator function of an event* $A$ *then* $\mathbb{E}(Y) = \mathbb{P}(A)$.

**Exercise 1.11.** *Calculate the expectation for the uniform random variable* $U$ *from Example 1.7.*

**Exercise 1.12.** *Prove Proposition 1.29. Hint: it is easiest to start with the case of positive discrete random variables and try to extend the results from there.*

**Exercise 1.13.** *A random variable* $X$ *takes values* $1, 2$ *and* $3$ *with* $\mathbb{P}(X = n) = cn^2$ *for* $n = 1, 2, 3$. *Find: (i) the value of the constant* $c$*; (ii)* $\mathbb{E}(X)$*; (iii)* $\mathbb{E}(1/X)$.

**Exercise 1.14.** *Show that if* $Y \sim \mathrm{Exp}(\lambda)$ *then* $\mathbb{E}(\exp(tY)) = \frac{\lambda}{\lambda - t}$ *for* $t < \lambda$.

**Exercise 1.15.** *Determine if the following are square integrable, and if so, compute their variance: (i) the discrete random variable* $X$ *with* $\mathbb{P}(X = n) = c/n^2$ *for all* $n \geq 1$ *and* $1/c = \sum_{n=1}^{\infty} 1/n^2$*; (ii) the exponential random variable* $Y \sim \mathrm{Exp}(\lambda)$*?*

**Exercise 1.16.** *Prove Proposition 1.35. Come up with an example to show that (iii) does not necessarily hold when* $X$ *and* $Y$ *are not independent.*

**Exercise 1.17.** *Consider three* ***independent*** *random variables* $X_1, X_2, X_3$ *with* $\mathbb{P}(X_i = \pm 1) = 1/2$ *for* $i = 1, 2$ *and* $\mathbb{P}(X_3 = 0) = \mathbb{P}(X_3 = 1) = 1/2$. *Letting* $Z_1 = X_1 X_3$ *and* $Z_2 = X_2 X_3$*, show that* $\mathbb{E}(Z_1 Z_2) = \mathbb{E}(Z_1)\mathbb{E}(Z_2)$. *What is the correlation between* $Z_1$ *and* $Z_2$*? Are* $Z_1, Z_2$ *independent? Hint: consider the event* $\mathbb{P}(Z_1 = 0 \cap Z_2 = 0)$.

**Solutions to exercises from §1**

1. We need to check that $\mathcal{F}_1 \cap \mathcal{F}_2$ satisfies conditions (i)-(iii) of the $\sigma$-field definition. For (i), since $\mathcal{F}_1$ and $\mathcal{F}_2$ are both $\sigma$-fields we have $\emptyset \in \mathcal{F}_1$ and $\emptyset \in \mathcal{F}_2$. Therefore, $\emptyset \in \mathcal{F}_1 \cap \mathcal{F}_2$. For (ii), suppose that $A \in \mathcal{F}_1 \cap \mathcal{F}_2$. This means that $A \in \mathcal{F}_1 \Rightarrow A^c \in \mathcal{F}_1$ and also that $A \in \mathcal{F}_2 \Rightarrow A^c \in \mathcal{F}_2$ (since $\mathcal{F}_1$, $\mathcal{F}_2$ are both $\sigma$-fields). Consequently $A^c \in \mathcal{F}_1 \cap \mathcal{F}_2$. Finally, if $(A_i)_{i\geq 1}$ are a sequence of subsets of $\Omega$ with $A_i \in \mathcal{F}_1 \cap \mathcal{F}_2$ for each $i$, we have $(A_i)_{i\geq 1} \in \mathcal{F}_1$ and $(A_i)_{i\geq 1} \in \mathcal{F}_2$. As with the previous points, this implies that $\cup_i A_i \in \mathcal{F}_1$ and $\cup_i A_i \in \mathcal{F}_2 \Rightarrow \cup_i A_i \in \mathcal{F}_1 \cap \mathcal{F}_2$, which means that $\mathcal{F}_1 \cap \mathcal{F}_2$ satisfies (iii).

2. Define $\mathcal{F}_1 = \{\emptyset, \{\omega_1,\omega_2,\omega_3\}, \{\omega_1\}, \{\omega_2,\omega_3\}\}$ and $\mathcal{F}_2 = \{\emptyset, \{\omega_1,\omega_2,\omega_3\}, \{\omega_2\}, \{\omega_1,\omega_3\}\}$. It is easy to check that $\mathcal{F}_1$ and $\mathcal{F}_2$ are both $\sigma$-fields. On the other hand, $\{\omega_1\} \in \mathcal{F}_1 \cup \mathcal{F}_2$ and $\{\omega_2\} \in \mathcal{F}_1 \cup \mathcal{F}_2$ but $\{\omega_1\} \cup \{\omega_2\} = \{\omega_1, \omega_2\} \notin \mathcal{F}_1 \cup \mathcal{F}_2$. So $\mathcal{F}_1 \cup \mathcal{F}_2$ is not a $\sigma$-field.

3. Suppose that $\mathcal{C}$ is a finite collection $A_1, \dots, A_n \subset \Omega$. Then

$$\sigma(\mathcal{C}) = \{\cup_{i\in I} A_i \bigcup \cup_{j\in J} A_j^c, \cap_{i\in I} A_i \bigcap \cap_{j\in J} A_j^c\}_{I,J}$$

where $I, J$ ranges over all disjoint subsets of $\{1,\dots,n\}$.

4. Recall that $\mathcal{B}(\mathbb{R})$ is defined to be the $\sigma$-field $\sigma(\mathcal{C})$ generated by the collection of intervals $\mathcal{C} := \{(a,b); -\infty \leq a < b \leq \infty\}$. To show that this is equal to $\sigma(\tilde{\mathcal{C}})$ for $\tilde{\mathcal{C}} = \{(-\infty, c]; c \in \mathbb{R}\}$ we need to show that (i) $\tilde{\mathcal{C}} \subset \sigma(\mathcal{C})$ and (ii) $\mathcal{C} \subset \sigma(\tilde{\mathcal{C}})$. Indeed **if we can show this** then (i) implies that $\sigma(\mathcal{C}) \supset \sigma(\tilde{\mathcal{C}})$ (remember that $\sigma(\tilde{\mathcal{C}})$ is by definition the smallest $\sigma$-field containing $\tilde{\mathcal{C}}$ so for any $\sigma$-field $\mathcal{F}$ containing $\tilde{\mathcal{C}}$ we must have $\mathcal{F} \supset \sigma(\tilde{\mathcal{C}})$) and similarly (ii) implies that $\sigma(\tilde{\mathcal{C}}) \supset \sigma(\mathcal{C})$. This means that $\sigma(\mathcal{C}) = \sigma(\tilde{\mathcal{C}})$. It is obvious that (i) holds since by definition elements of $\tilde{\mathcal{C}}$ are also elements of $\mathcal{C}$. (ii) is slightly more tricky, but for any $(a,b) \in \mathcal{C}$ we can write $(a,b) = ([b,\infty) \cup (-\infty, a])^c$. Note that $(-\infty, a] \in \tilde{\mathcal{C}}$ by definition, and that $[b,\infty) = (\cup_n(-\infty, b - 2^{-n}])^c$ must also be in $\sigma(\tilde{\mathcal{C}})$ since it is the complement of the union of elements of $\tilde{\mathcal{C}}$. This implies that $(a,b)$ is the complement of a union of two events in $\sigma(\tilde{\mathcal{C}})$ and therefore $(a,b) \in \sigma(\tilde{\mathcal{C}})$ (since $\sigma(\tilde{\mathcal{C}})$ is a $\sigma$-field). Hence (i) and (ii) have been shown, and we are done.

5. Observe that we can write $A \cup B = (A \cap B^c) \cup B$ and also $A = (A \cap B^c) \cup (A \cap B)$ as disjoint unions of events. The latter tells us that $\mathbb{P}(A) = \mathbb{P}(A \cap B) + \mathbb{P}(A \cap B^c)$ so that $\mathbb{P}(A \cap B^c) = \mathbb{P}(A) - \mathbb{P}(A \cap B)$ while the former says that $\mathbb{P}(A \cup B) = \mathbb{P}(A \cap B^c) + \mathbb{P}(B)$. Putting these together gives the first result. For the second result, writing $A^c \cap (B \cup C) = (A^c \cap B) \cup (A^c \cap C)$ and applying the first result with $A = A^c \cap B$ and $B = A^c \cap C$, we get that

$$\mathbb{P}(A^c \cap (B \cup C)) = \mathbb{P}(A^c \cap B) + \mathbb{P}(A^c \cap C) - \mathbb{P}(A^c \cap B \cap C).$$

Writing $\mathbb{P}(E \cap A^c) = \mathbb{P}(E) - \mathbb{P}(E \cap A)$ with $E = B$, $E = C$ and $E = B \cap C$ and then substituting into the above then provides the desired equality.

For the final part of the exercise, let $\Omega = \{1, \dots, 500\}$, $\mathcal{F} = \mathcal{P}(\Omega)$, and set $\mathbb{P}(\omega) = 1/500$ for each element of $\Omega$. Then using the above with $A = \{7x : x \in \mathbb{Z}, 0 \leq 7x \leq 500\}$, $B = \{5x : x \in \mathbb{Z}, 0 \leq 5x \leq 500\}$ and $C = \{3x : x \in \mathbb{Z}, 0 \leq 3x \leq 500\}$ gives $\mathbb{P}(B) = 100/500$ and $\mathbb{P}(C) = \lfloor 500/3 \rfloor/500 = 166/500 =$, while $\mathbb{P}(A \cap B) = \lfloor 500/35 \rfloor/500 = 14/500$, $\mathbb{P}(A \cap C) = \lfloor 500/21 \rfloor/500 = 23/500$, $\mathbb{P}(B \cap C) = \lfloor 500/15 \rfloor/500 = 33/500$ and $\mathbb{P}(A \cap B \cap C) = \lfloor 500/105 \rfloor/500 = 4/500$. The answer is then $500 * (\mathbb{P}(A^c \cap (B \cup C))) = 100 - 14 + 166 - 23 + 4 - 33 = 200$.

6. Write $F$ for the event that the component is faulty, and $A, B, C$ for the events that the component is drawn from $A, B, C$ respectively. (i) is asking for the probability of $F$ which is equal to $\mathbb{P}(F \cap A) + \mathbb{P}(F \cap B) + \mathbb{P}(F \cap C) = \mathbb{P}(F|A)\mathbb{P}(A) + \mathbb{P}(F|B)\mathbb{P}(B) + \mathbb{P}(F|C)\mathbb{P}(C) = 0.1 * (1/3) + 0.2 * (1/3) + 0.3 * (1/3) = 0.6 * (1/3) = 0.2$. (ii) is asking for $\mathbb{P}(A|F) = \mathbb{P}(A \cap F)/\mathbb{P}(F) = \mathbb{P}(F|A)\mathbb{P}(A)/\mathbb{P}(F) = 0.1 * (1/3)/0.2 = 1/6$.

7. We work on $([0,1], \mathcal{B}([0,1]), \mathbb{P})$ where $\mathbb{P}((a,b)) = b-a$ for any open interval $(a,b)$, and are considering the random variable $U(\omega) = \omega$. Then if $x \in (0,1)$,

$$\mathbb{P}(U \in (-\infty, x]) = \mathbb{P}(\{\omega \in \Omega : U(\omega) \leq x\}) = \mathbb{P}(\{\omega \in [0,1] : \omega \leq x\}) = \mathbb{P}([0,x]).$$

Note that $\mathbb{P}(\{x\}) = 0$ since choosing any interval $[0,1] \supset (a,b) \ni x$ gives $b - a = \mathbb{P}((a,b)) = \{x\} \cup (a,x) \cup (x,b)$, which is a disjoint union so $b - a = \mathbb{P}(\{x\}) + (x-a) + (b-x) = \mathbb{P}(\{x\}) + b - a$. Therefore $\mathbb{P}([0,x]) = \mathbb{P}(\{0\}) + \mathbb{P}(\{x\}) + \mathbb{P}((0,x)) = x$. If $x \leq 0$ it is clear that $\mathbb{P}(U \in (-\infty, x]) = 0$ and if $x \geq 1$ then $\mathbb{P}(U \in (-\infty, x]) = 1$. So the distribution function $F_U$ is just equal to 0 for $x \leq 0$ and $\min(x, 1)$ for $x \geq 0$. Note that we therefore have $F_U(x) = \int_{-\infty}^{x} \mathbf{1}_{y \in [0,1]}\, dy$ for every $x$ and so $U$ has a probability distribution function $p_U(y) = \mathbf{1}_{y \in [0,1]}$.

8. $\sigma(Y)$ is by definition the smallest $\sigma$-field with respect to which $Y = \mathbf{1}_{\{2,4,6\}}$ is measurable. Observe that $Y$ is measurable with respect to $\mathcal{F} := \{\emptyset, \Omega, \{2,4,6\}, \{1,3,5\}\}$ which **is** a $\sigma$-field. Moreover, any $\sigma$-field with respect to which $Y$ is measurable has to contain the event $Y^{-1}((-\infty, 1/2]) = \{\omega \in \Omega : Y(\omega) \leq 1/2\} = \{1,3,5\}$ and since it is a $\sigma$-field, must therefore also contain the complement $\{2,4,6\}$ and the sets $\emptyset, \Omega$. So $\mathcal{F}$ is the smallest possible $\sigma$-field with respect to which $Y$ is measurable $\Rightarrow \mathcal{F} = \sigma(Y)$.

9. If $X$ and $Y$ are independent and $x \in \mathcal{X}$, $y \in \mathcal{Y}$ then $\{X = x\} \in \sigma(X)$ and $\{Y = y\} \in \sigma(Y)$ so that $\mathbb{P}(\{X = x\} \cap \{Y = y\}) = \mathbb{P}(\{X = x\})\mathbb{P}(\{Y = y\})$. Conversely, if this holds for all $x \in \mathcal{X}, y \in \mathcal{Y}$ we can write any $A \in \sigma(X), B \in \sigma(Y)$ as $A = \cup_{x \in \mathcal{X}_1}\{X = x\}$ where the union is over some subset $\mathcal{X}_1 \subset \mathcal{X}$ and $B = \cup_{y \in \mathcal{Y}_1}\{Y = y\}$ where the union is over some subset $\mathcal{Y}_1 \subset \mathcal{Y}$. Then

$$\begin{aligned}
\mathbb{P}(A \cap B) &= \mathbb{P}(\cup_{x \in \mathcal{X}_1, y \in \mathcal{Y}_1}\{X = x\} \cap \{Y = y\}) \\
&= \sum_{x \in \mathcal{X}_1, y \in \mathcal{Y}_1} \mathbb{P}(X = x)\mathbb{P}(Y = y) \\
&= (\sum_{x \in \mathcal{X}_1} \mathbb{P}(X = x))(\sum_{y \in \mathcal{Y}_1} \mathbb{P}(Y = y)) \\
&= \mathbb{P}(A)\mathbb{P}(B)
\end{aligned}$$

$\Rightarrow X$ and $Y$ are independent. In the example of a dice roll $X(\omega) = \omega$ and $Y(\omega) = \{1,3,5\}$ are **not** independent, because $\mathbb{P}(\{Y = 1\} \cap \{X = 2\}) = 0 \neq \mathbb{P}(Y = 1) \times \mathbb{P}(X = 2)$.

10. If $Y = \mathbf{1}_A$ then $Y$ is a discrete random variable taking value 1 with probability $\mathbb{P}(A)$ and 0 with probability $1 - \mathbb{P}(A)$. So $\mathbb{E}(Y) = 1^*\mathbb{P}(A) + 0^*(1 - \mathbb{P}(A)) = \mathbb{P}(A)$.

11. We saw above that the probability density function of a uniform random variable is given by $p_u(x) = \mathbf{1}_{x \in [0,1]}$. Therefore, $\mathbb{E}(U) = \int_{-\infty}^{\infty} x \mathbf{1}_{x \in [0,1]}\, dx = \int_0^1 x\, dx = 1/2$.

12. (i) If $\mathbb{P}(X \geq 0) = 1$ then by definition $\mathbb{E}(X)$ is an increasing limit of $\mathbb{E}(X_n)$ where the $X_n$ are discrete and positive. In the discrete case, since $\mathbb{E}(X) = \sum_x x\mathbb{P}(X = x)$ it is therefore immediate that $\mathbb{P}(X \geq 0) = 1 \Rightarrow \mathbb{E}(X) \geq 0$. The proofs of properties (ii)-(v) are similar. For (vi), we have $|\mathbb{E}(X)| = |\mathbb{E}(X^+) - \mathbb{E}(X^-)| \leq |\mathbb{E}(X^+)| + |\mathbb{E}(X^-)|$ by the triangle inequality for real numbers. But $X^+$ and $X^-$ are both positive with probability one, so by (i) of this exercise, $|\mathbb{E}(X^\pm)| = \mathbb{E}(X^\pm)$ and $\mathbb{E}(|X|) = \mathbb{E}(X^+\mathbf{1}_{X \geq 0} + X^-\mathbf{1}_{\{X<0\}}) = E(X^+) + \mathbb{E}(X^-)$. Putting these together gives $|\mathbb{E}(X)| \leq \mathbb{E}(|X|)$ as required.

13. (i) It must be that $\mathbb{P}(X = 1) + \mathbb{P}(X = 2) + \mathbb{P}(X = 3) = 1$, so $c + 4c + 9c = 1 \Rightarrow c = 1/14$. (ii) $\mathbb{E}(X) = 1^*\mathbb{P}(X = 1) + 2^*\mathbb{P}(X = 2) + 3^*\mathbb{P}(X = 3) = (1/14)^*(1 + 8 + 27) = 36/14$. (iii) $\mathbb{E}(1/X) = 1^*\mathbb{P}(X = 1) + (1/2)^*\mathbb{P}(X = 2) + (1/3)^*\mathbb{P}(X = 3) = (1/14)^*(1 + 2 + 3) = 6/14$.

14. By definition, $p_Y(y) = \lambda\, \mathrm{e}^{-\lambda y}\, \mathbf{1}_{y \geq 0}$ so $\mathbb{E}(\exp(tY)) = \int_0^\infty \lambda e^{-\lambda y} e^{ty}\, dy$ which is infinite if $t > \lambda$ and equal to $\lambda/(\lambda - t)$ if $t < \lambda$.

15. (i) $X$ is square integrable iff $\sum_{n=1}^\infty n^2 \mathbb{P}(X = n)$ is finite. When $\mathbb{P}(X = n) = c/n^2$ this is not the case. (ii) On the other hand, $Y$ is square integrable: all we have to check is that $\int_0^\infty y^2 \lambda e^{-\lambda y} < \infty$ which is true, it is equal to $1/\lambda^2$.

16. (i) $\mathrm{var}(X)$ is always $\geq 0$, since it is the expectation of a square, i.e., a positive random variable. If $\mathrm{var}(X) = 0$ then by definition $\mathbb{E}((X - \mathbb{E}(X))^2) = 0$. But suppose $Y$ is a positive random variable with $\mathbb{E}(Y) = 0$. Then recalling the definition $Y_n = \min(n, 2^{-n}\lfloor 2^n Y \rfloor)$ we have that $0 \leq \mathbb{E}(Y_n) \leq \mathbb{E}(Y)$ for each $n$ and so $\mathbb{E}(Y_n) = 0$ for all $n$. Moreover, for every fixed $n$, $Y_n$ is discrete taking values in $\{0, 2^{-n}, 2.2^{-n}, 3.2^{-n}, \ldots, n\}$, so it must be that $\mathbb{P}(Y_n = x) = 0$ for $x \neq 0$ and therefore $\mathbb{P}(Y_n = 0) = 1$. Since this holds for all $n$, we see that $Y = 0$ with probability one. Going back to the setting of the question, we can deduce that $(X - \mathbb{E}(X))^2 = 0$ with probability one, and therefore $X$ is constant. Property (ii) holds since

$$\mathrm{var}(aX) = \mathbb{E}((aX - \mathbb{E}(aX))^2) = \mathbb{E}((aX - a\mathbb{E}(X))^2) = \mathbb{E}(a^2(X - \mathbb{E}(X))^2) = a^2\mathbb{E}((X - \mathbb{E}(X))^2)$$

which is exactly $a^2\,\mathrm{var}(X)$. For (iii), if $X$ and $Y$ are independent, then

$$\begin{aligned}
\mathrm{var}(X+Y) &= \mathbb{E}((X + Y - \mathbb{E}(X+Y))^2) \\
&= \mathbb{E}((X + Y - \mathbb{E}(X) - \mathbb{E}(Y))^2) \\
&= \mathbb{E}((X - \mathbb{E}(X))^2 + (Y - \mathbb{E}(Y))^2 - 2(X - \mathbb{E}(X))(Y - \mathbb{E}(Y))) \\
&= \mathrm{var}(X) + \mathrm{var}(Y) - 2\,\mathrm{cov}(X, Y) \\
&= \mathrm{var}(X) + \mathrm{var}(Y),
\end{aligned}$$

as required.

Finally, let $X$ be any non-constant random variable and $Y = -X$. Then $\mathrm{var}(Y + X) = 0$ but $\mathrm{var}(X) + \mathrm{var}(Y) = 2\,\mathrm{var}(X) \neq 0$.

17. $Z_1 Z_2$ is equal to 0 if $X_3 = 0$, 1 if $X_1 = X_2$ and $X_3 = 1$, and $-1$ if $X_1 \neq X_2$ and $X_3 = 1/2$. So $\mathbb{P}(Z_1 Z_2 = 1) = 1/4 = \mathbb{P}(Z_1 Z_2 = -1)$ and $\mathbb{P}(Z_1 Z_2 = 0) = 1/2$. This means that $\mathbb{E}(Z_1 Z_2) = 0$ while $\mathbb{E}(Z_1), \mathbb{E}(Z_2)$ are also 0 by similar reasoning. $\mathrm{cov}(Z_1, Z_2)$ is therefore equal to $\mathbb{E}(Z_1 Z_2) - \mathbb{E}(Z_1)\mathbb{E}(Z_2) = 0$, meaning that the correlation between $Z_1$ and $Z_2$ is also 0. $Z_1$ and $Z_2$ are not independent however, since $\mathbb{P}(\{Z_1 = 0\} \cap \{Z_2 = 0\}) = \mathbb{P}(X_3 = 0) = 1/2$ but $\mathbb{P}(Z_1 = 0)\mathbb{P}(Z_2 = 0) = 1/2 * 1/2 = 1/4$ (and if they were independent then these should be equal).

# 2 Normal random variables and convergence

Complementary reading: [1, Appendix 1].

## 2.1 Convergence of random variables

**Definition 2.1** (Almost sure convergence)**.** *Let $X$ and $X_0, X_1, X_2, \ldots$ be random variables on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. We say that $X_n$ converges to $X$* **almost surely***, written $X_n \overset{a.s.}{\to} X$, if*

$$\mathbb{P}\left(\left\{\omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)\right\}\right) = 1. \tag{38}$$

Another way to say the same thing is that if $N_\varepsilon$ (random) is the smallest integer such that

$$|X_n - X| \leq \varepsilon, \text{ for all } n \geq N_\varepsilon, \tag{39}$$

then $X_n \to X$ a.s. if and only if $\mathbb{P}\left(N_\varepsilon < \infty \text{ for all } \varepsilon > 0\right) = 1$.

**Example 2.2** (Strong law of large numbers)**.** *Suppose that $X_1, X_2, \ldots$ are a sequence of independent, identically distributed and integrable random variables, with mean $\mu$. Then*

$$\bar{S}_n := \frac{\sum_{i=1}^n X_i}{n} \xrightarrow{a.s.} \mu \tag{40}$$

*as $n \to \infty$.*

**Definition 2.3** (Convergence in probability)**.** *Let $X$ and $X_0, X_1, X_2, \ldots$ be random variables on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. We say that $X_n$ converges to $X$* **in probability***, written $X_n \overset{\mathbb{P}}{\to} X$, if*

$$\lim_{n\to\infty} \mathbb{P}\left(|X_n - X| > \varepsilon\right) = 0 \textit{ for all } \varepsilon > 0. \tag{41}$$

**Definition 2.4** (Convergence in $L^q$)**.** *Let $X$ and $X_0, X_1, X_2, \ldots$ be random variables on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. For $q \geq 1$, we say $X_n$ converges to $X$* **in** $L^q$*, if*

$$\lim_{n\to\infty} \mathbb{E}\left[|X_n - X|^q\right] = 0. \tag{42}$$

Since by the triangle inequality

$$|\mathbb{E}(|X_n|) - \mathbb{E}(|X|)| \leq \mathbb{E}(|X_n - X|), \tag{43}$$

$X_n \to X$ in $L^1$ implies that $\mathbb{E}(|X_n|) \to \mathbb{E}(|X|)$. The same holds with $q$ in place of 1 for every $q \geq 1$.

**Definition 2.5** (Convergence in distribution/law)**.** *Let $X$ and $X_0, X_1, X_2, \ldots$ be random variables (not necessarily on the same probability space) with cumulative distribution functions*

$$F(x) = \mathbb{P}(X \leq x) \textit{ and } F_n(x) = \mathbb{P}(X_n \leq x). \tag{44}$$

*We say $X_n$ converges to $X$* **in distribution***, written $X_n \overset{(d)}{\to} X$, if*

$$\lim_{n\to\infty} F_n(x) = F(x) \textit{ for all } x \textit{ at which } F \textit{ is continuous.} \tag{45}$$

**Example 2.6.** *For every $n \geq 1$, let $X_n$ be the continuous random variable with cumulative distribution function $F_n(x) = \mathbb{P}(X_n \leq x) = 1 - (1 - n^{-1}x)^n$. Then $F_n(x) \to 1 - e^{-x}$ for every $x \in \mathbb{R}$. But $F(x) = 1 - \mathrm{e}^{-x} = \int_0^x e^{-y}\,dy$ is the cumulative distribution function of an exponential random variable $X \sim \mathrm{Exp}(1)$. So $X_n \to X$ in distribution as $n \to \infty$.*

**Lemma 2.7.** *The following implications hold:*

- *$X_n \to X$ a.s. implies that $X_n \overset{\mathbb{P}}{\to} X$.*
- *$X_n \to X$ in $L^q$ implies that $X_n \overset{\mathbb{P}}{\to} X$.*
- *$X_n \overset{\mathbb{P}}{\to} X$ implies that $X_n \overset{(d)}{\to} X$.*

*Proof.* Let us show the implications in order.

- To prove the first implication, note that $\mathbb{P}(|X_n - X| > \varepsilon) \leq \mathbb{P}(n < N_\varepsilon)$, so

$$\lim_{n\to\infty} \mathbb{P}\left(|X_n - X| > \varepsilon\right) \leq \lim_{n\to\infty} \mathbb{P}(n < N_\varepsilon) = \mathbb{P}(N_\varepsilon = \infty). \tag{46}$$

- The second follows from Markov's inequality since $\mathbb{P}(|X_n - X| > \varepsilon) \leq \varepsilon^{-q}\mathbb{E}(|X_n - X|^q)$.
- Suppose $x \in \mathbb{R}$ is such that $F(x) := \mathbb{P}(X \leq x)$ is continuous at $x$. Then for any $\varepsilon > 0$, it is straightforward to check that

$$\mathbb{P}(X_n \leq x) \in [\mathbb{P}(X \leq x - \varepsilon) + \mathbb{P}(|X_n - X| > \varepsilon), \mathbb{P}(X \leq x + \varepsilon) + \mathbb{P}(|X_n - X| > \varepsilon)]. \tag{47}$$

By continuity of $F$ at $x$ and the assumption that $\mathbb{P}(|X_n - X| > \varepsilon) \to 0$ as $\varepsilon \to 0$, it follows that the left and right end points of the interval above converge to $\mathbb{P}(X \leq x)$ as $n \to \infty$.

□

**Example 2.8.** *Suppose that $X_n = 2^n$ with probability $2^{-n}$ for every $n$, $X_n = 2^{-n}$ otherwise. Let $X = 0$ with probability one. Then*

$$F_{X_n}(x) = \begin{cases} 0 & \text{if } x < 2^{-n} \\ 1 - 2^{-n} & \text{if } x \in [2^{-n}, 2^n) \\ 1 & \text{if } x \geq 2^{-n} \end{cases} \tag{48}$$

*which converges to $\mathbf{1}_{\{x \geq 0\}}$ for all $x \neq 0$. Since $0$ is not a continuity point of $F_X$, it follows that $X_n \to X$ in distribution as $n \to \infty$. Note however that $\mathbb{E}(X_n) \geq 1$ for each $n$, while $\mathbb{E}(X) = 0$.*

## Convergence of expectations

Let us state here without proof some useful results about when expectations converge.

**Lemma 2.9** (Fatou's lemma)**.** *Suppose that $X_1, \ldots$ are a sequence of non-negative random variables and $X_n \to X$ almost surely. Then*

$$\mathbb{E}(X) \leq \liminf_{n\to\infty} \mathbb{E}(X_n). \tag{49}$$

**Theorem 2.10** (Monotone convergence theorem)**.** *Suppose that $X_1, \ldots$ are a sequence of non-negative random variables that are almost surely increasing in $n$ and converge to $X$ almost surely as $n \to \infty$. Then*

$$\mathbb{E}(X_n) \uparrow \mathbb{E}(X) \tag{50}$$

*as $n \to \infty$.*

**Theorem 2.11** (Dominated convergence theorem)**.** *Suppose that $X_1, \ldots$ are a sequence of random variables that converge to $X$ almost surely as $n \to \infty$. Suppose further that for some non-negative integrable random variable $Z$, $|X_n| \leq Z$ almost surely for all $n$. Then*

$$\mathbb{E}(X_n) \to \mathbb{E}(X) \tag{51}$$

*as $n \to \infty$.*

### Exercises from §2.1

**Exercise 2.1.** *Take the set-up of Example 2.2 and assume in addition that the* $(X_i)_{i\geq 1}$ *have finite variances. Use Markov's inequality from §1 to show that* $\bar{S}_n \xrightarrow{\mathbb{P}} \mu$ *as* $n \to \infty$*.*

**Exercise 2.2.** *If* $X_n \sim \mathrm{Exp}(n)$ *for every* $n$*, show that* $X_n \xrightarrow{(d)} 0$ *as* $n \to \infty$*.*

## 2.2 The normal distribution and the central limit theorem

**Definition 2.12** (Normal distribution)**.** *A random variable* $X$ *has the* **normal (or Gaussian) distribution** *with* **mean** $\mu$ *and* **variance** $\sigma^2$ *if it is a continuous random variable with probability density function*

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}, \text{ for } x \in \mathbb{R}. \tag{52}$$

*It is denoted by* $X \sim \mathcal{N}(\mu, \sigma^2)$*.*

Note that $\int_{-\infty}^{\infty} p(x)\, dx = 1$ (you can take this for granted, or see Wikipedia for a proof!)

The case $\mu = 0$, $\sigma^2 = 1$ is the **standard normal** distribution $Z \sim \mathcal{N}(0,1)$. The density of the standard normal is usually written as $\phi$, and the **cumulative distribution function** is

$$\Phi(x) = \Pr(Z \leq x) = \int_{-\infty}^{x} \phi(y)\mathrm{d}y. \tag{53}$$

If $X \sim \mathcal{N}(\mu, \sigma^2)$, then $\alpha + \beta X \sim \mathcal{N}(\alpha + \beta\mu, \beta^2\sigma^2)$. In particular, if $X \sim \mathcal{N}(\mu, \sigma^2)$, then

$$\frac{X-\mu}{\sigma} \sim \mathcal{N}(0,1). \tag{54}$$

This implies that if $X \sim \mathcal{N}(\mu, \sigma^2)$ then $\mathbb{E}(X) = \mu$ and $\mathrm{var}(X) = \sigma^2$; because

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x e^{-x^2/2} = 0 \text{ and } \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^2 e^{-x^2/2}\, dx = 1$$

(the first equality follows by symmetry and the second by integration by parts with $u = -x$, $v' = -xe^{-x^2/2}$, using that $\int_{-\infty}^{\infty} e^{-x^2/2} dx = \sqrt{2\pi}$).

The normal random variable is fundamental in virtually all applications of probability theory; especially because of the **central limit theorem** (see below). Its probability density function is the familiar bell-shaped curve: this has peak at $\mu$ and is more spread out as $\sigma^2$ increases.

**Definition 2.13.** *A random vector* $\mathbf{X} = (X_1, \cdots, X_n)$ *is* ***normally distributed****, also referred to as a* ***Gaussian random vector****, if and only if* $a_1X_1 + \cdots + a_nX_n$ *has a normal distribution for any* $(a_1, \ldots, a_n) \in \mathbb{R}^n$*. If*

$$\mu := (\mathbb{E}(X_1), \cdots, \mathbb{E}(X_n)) \text{ and } \Sigma \text{ is an } n \times n \text{ matrix with } \Sigma_{ij} = \mathrm{cov}(X_i, X_j)\ \forall 1 \leq i,j \leq n \tag{55}$$

*it is said that* $\mathbf{X} \sim (\mu, \Sigma)$*. When* $\Sigma$ *is positive definite,* $\mathbf{X}$ *has probability density*

$$p_{\mathbf{X}}(\mathbf{x}) = \frac{\exp(-\frac{1}{2}(\mathbf{x}-\mu)^T \Sigma^{-1}(\mathbf{x}-\mu)}{\sqrt{(2\pi)^n |\det(\Sigma)|}}\ ;\ \mathbf{x} = (x_1, \ldots, x_n), \tag{56}$$

*meaning that for any* $B \subset \mathbb{R}^n$*,*

$$\mathbb{P}(\mathbf{X} \in B) = \int_B p_{\mathbf{X}}(x_1, \ldots, x_n)\, dx_1 \ldots dx_n \tag{57}$$

**The central limit theorem**

Let $X_1, X_2, \ldots$ be **independent, identically distributed** (i.i.d.) random variables on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Set $S_0 = 0$ and $S_n = \sum_{i=1}^n X_i$ for $n \geq 1$.

**Theorem 2.14** (Central Limit Theorem (CLT))**. The central limit theorem** *says that if* $\mathbb{E}(X_i) = \mu$ *and* $\text{var}(X_i) = \sigma^2 \in (0, \infty)$*, then*

$$\frac{S_n - n\mu}{\sqrt{n\sigma^2}} \stackrel{(d)}{\to} \mathcal{N}(0,1). \tag{58}$$

In other words,

$$\lim_{n\to\infty} \mathbb{P}\left[\frac{S_n - n\mu}{\sqrt{n\sigma^2}} \leq z\right] = \Phi(z) \text{ for all } z \in \mathbb{R}. \tag{59}$$

*Proof.* Omitted. □

## Exercises from §2.2

**Exercise 2.3.** *Let* $Z \sim \mathcal{N}(0,1)$*. Show that*

$$M_Z(t) = \mathbb{E}\left(\mathrm{e}^{tZ}\right) = \mathrm{e}^{t^2/2} \tag{60}$$

*for any* $t \in \mathbb{R}$*.*

**Exercise 2.4.** *Suppose that* $Y_1, Y_2, \ldots, Y_n$ *are independent with* $Y_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$ *for each* $i$*. Show that* $\sum_{i=1}^n Y_i \sim \mathcal{N}(\sum_i \mu_i, \sum_i \sigma_i^2)$*.*

**Exercise 2.5.** *Suppose that* $Z_1, \ldots, Z_n$ *are i.i.d. standard normal random variables. Write* $\mathbf{Z} = (Z_1, \ldots, Z_n)$ *and suppose that* $A$ *is an* $n \times n$ *matrix and* $\mathbf{a}$ *is a vector of length* $n$*. Show that* $\mathbf{X} = A\mathbf{Z} + \mathbf{a}$ *is a normally distributed vector, and find its mean and covariance matrix.*

**Exercise 2.6.** *Explain why the Theorem 2.14 holds when* $X_i \sim \mathcal{N}(\mu, \sigma^2)$ *for each* $i$*. Hint: you should use Exercise 2.4 .*

**Solutions to exercises from §2**

1. Fix $\varepsilon > 0$. Then by Markov's inequality

$$\mathbb{P}(|\bar{S}_n - \mu| > \varepsilon) = \mathbb{P}(|\bar{S}_n - \mathbb{E}(\bar{S}_n))| > \varepsilon) \leq \frac{\operatorname{var}(\bar{S}_n)}{\varepsilon^2}$$

by Markov's inequality. But

$$\operatorname{var}(\bar{S}_n) = \operatorname{var}(\frac{1}{n}\sum_{i=1}^n X_i) = \frac{1}{n^2}\sum_{i=1}^n \operatorname{var}(X_i) = \frac{1}{n}\operatorname{var}(X_1),$$

using the basic properties of variance and the fact that the $(X_i)_{i\geq 1}$ are i.i.d. This goes to 0 as $n \to \infty$, so indeed $\mathbb{P}(|\bar{S}_n - \mu| > \varepsilon) \to 0$ as $n \to \infty$. Since this holds for any $\varepsilon$, we have shown that $\bar{S}_n$ tends to $\mu$ in probability.

2. We need to show that if $Y_n \sim \operatorname{Exp}(n)$ and $Y \equiv 0$, $F_{Y_n}(x) \to F_Y(x)$ for all $x$ such that $F_Y$ is continuous at $x$. Note that $F_Y(x) = \mathbb{P}(Y \leq x)$ is equal to 0 if $x < 0$ and 1 if $x > 0$, so is continuous everywhere except at 0. Therefore we need to show that $\mathbb{P}(Y_n \leq x) \to 0$ as $n \to \infty$ for any $x > 0$ and $\mathbb{P}(Y_n \leq x) \to 1$ as $n \to \infty$ for any $x > 0$. The first of these points is obvious since $\mathbb{P}(Y_n < 0) = 0$ for all $n$. For the second point we can just calculate $\mathbb{P}(Y_n \leq x) = \int_0^x ne^{-ny}\,dy = 1 - \mathrm{e}^{-nx}$. This does tend to 1 as $n \to \infty$, so we are done.

20. The key point here is that we know, for any $\mu, \sigma \in \mathbb{R}$, that $(2\pi\sigma^2)^{-1/2}\int_{-\infty}^{\infty}\exp(-(x-\mu)^2/(2\sigma^2))\,dx = 1$, because it is the integral of a probability density function.

We wish to calculate for $Z \sim \mathcal{N}(0,1)$:

$$\mathbb{E}(e^{tZ}) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{tx}e^{-x^2/2\,dx}.$$

The trick is that $e^{-(x-t)^2/2} = e^{tx}e^{-x^2/2}e^{-t^2/2}$ so that the above is equal to

$$e^{t^2/2}\left(\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-(x-t)^2/2}\,dx\right) = e^{t^2/2} \times 1 = e^{t^2/2}$$

as required.

By applying a change of variables, we can extend this to show that when $Z \sim \mathcal{N}(0,\sigma^2)$, $\mathbb{E}(e^{tZ}) = e^{t^2\sigma^2/2}$ for any $t$.

3. By induction it suffices to consider the case $n = 2$. Moreover, by subtracting $\mu_1$ from $Y_1$ and $\mu_2$ from $Y_2$, it suffices to consider the case $\mu_1 = \mu_2 = 0$.

So suppose that $Y_1 \sim \mathcal{N}(0,\sigma_1^2)$ and $Y_2 \sim \mathcal{N}(0,\sigma_2^2)$. Then for any $x \in \mathbb{R}$,

$$\begin{aligned}\mathbb{P}(Y_1 + Y_2 \leq x) &= \iint_{\{y_1+y_2\leq x\}} \tfrac{1}{\sqrt{2\pi}\sigma_1}e^{-y_1^2/(2\sigma_1^2)}\tfrac{1}{\sqrt{2\pi}\sigma_2}e^{-y_2^2/(2\sigma_2^2)}dy_1dy_2 \\ &= \tfrac{1}{2\pi\sigma_1\sigma_2}\int_{-\infty}^{\infty}\int_{-\infty}^{x-y_2} e^{-y_1^2/(2\sigma_1^2)}e^{-y_2^2/(2\sigma_2^2)}\,dy_1dy_2 \\ &= \tfrac{1}{2\pi\sigma_1\sigma_2}\int_{-\infty}^{\infty}\int_{-\infty}^{x} e^{-(u-y_2)^2/(2\sigma_1^2)}e^{-y_2^2/(2\sigma_2^2)}dudy_2\end{aligned}$$

where the last line follows by making the change of variables $u(y_1) = y_1 + y_2$. Next, we can switch the order of integration to rewrite this as

$$\frac{1}{2\pi\sigma_1\sigma_2}\int_{-\infty}^{x} e^{-u^2/(2\sigma_1^2)}\int_{-\infty}^{\infty} e^{-\frac{y_2^2}{2(\sigma_1^2\sigma_2^2/\sigma_1^2+\sigma_2^2)}}e^{\frac{uy_2}{\sigma_1^2}}\,dy_2du$$

where we know by the previous exercise that writing $\sigma^2 = (\sigma_1^2\sigma_2^2/\sigma_1^2+\sigma_2^2)$ we have $\frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty} e^{ty}e^{-y^2/2\sigma^2} = e^{t^2\sigma^2/2}$ for all $t$. Substituting this in, we see that

$$\int_{-\infty}^{\infty} e^{-\frac{y_2^2}{2(\sigma_1^2\sigma_2^2/\sigma_1^2+\sigma_2^2)}} e^{\frac{uy_2}{\sigma_1^2}} dy_2 = \sigma_1\sigma_2\sqrt{2\pi}\frac{1}{\sqrt{\sigma_1^2+\sigma_2^2}} e^{\frac{u^2}{2\sigma_1^4}\frac{\sigma_1^2\sigma_2^2}{\sigma_1^2+\sigma_2^2}}$$

and hence

$$\mathbb{P}(Y_1+Y_2 \leq x) = \frac{1}{\sqrt{2\pi(\sigma_1^2+\sigma_2^2)}}\int_{-\infty}^{x} e^{-\frac{u^2}{2}\frac{1}{\sigma_1^2}(1-\frac{\sigma_2^2}{\sigma_1^2+\sigma_2^2})} = \frac{1}{\sqrt{2\pi(\sigma_1^2+\sigma_2^2)}}\int_{-\infty}^{x} e^{-\frac{u^2}{2(\sigma_1^2+\sigma_2^2)}}.$$

This says that $Y_1 + Y_2 \sim \mathcal{N}(0, \sigma_1^2+\sigma_2^2)$, as required.

4. We first note that $\mathbf{Z} = (Z_1, \ldots, Z_n)$ is a Gaussian vector, by the previous exercise. Now let $\mathbf{X} = (X_1, \ldots, X_n)$. Recalling the definition of a Gaussian vector, we need to show that $Y := b_1X_1 + \cdots + b_nX_n$ is normally distributed for any $\mathbf{b} = (b_1, \ldots, b_n) \in \mathbb{R}^n$. But $Y = \mathbf{b}^T\mathbf{X}$ and $\mathbf{X} = A\mathbf{Z} + \mathbf{a}$ by definition, so

$$Y = \mathbf{b}^T(A\mathbf{Z}) + \mathbf{b}^T\mathbf{a} = (A\mathbf{b})^T\mathbf{Z} + \mathbf{b}^T\mathbf{a}$$

where $(\mathbf{Ab})^T\mathbf{Z}$ is normal since $\mathbf{Z}$ is multivariate Gaussian. Therefore $Y$ is a normal random variable plus a real number, meaning that it is also Gaussian.

By linearity of expectation (and since the mean vector of $\mathbf{Z}$ is $\mathbf{0}$ by definition) the mean vector of $\mathbf{X}$ is just equal to $\mathbf{a}$. To calculate the covariance matrix, observe (or check!) that adding a constant vector does **not** affect this, so

$$\text{cov}(X_i, X_j) = \text{cov}((A\mathbf{Z})_i, (A\mathbf{Z})_j) = \mathbb{E}((A\mathbf{Z})_i(A\mathbf{Z})_j) = \mathbb{E}((\sum_{k=1}^{n} A_{ik}Z_k)(\sum_{k=1}^{n} A_{jk}Z_k)),$$

where the penultimate equality follows since $\mathbb{E}((A\mathbf{Z})_i) = 0$ for $i = 1, \ldots, n$. Multiplying out the sums inside the expectation, and since $\mathbb{E}(Z_iZ_j) = \mathbb{E}(Z_i)\mathbb{E}(Z_j) = 0$ for $i \neq j$ by independence and $\mathbb{E}(Z_i^2) = 1$ for all $i$, this expectation is equal to $\sum_{k=1}^{n} A_{ik}A_{jk} = (A^TA)_{ij}$. In other words, the covariance matrix of $\mathbf{X}$ is equal to $A^TA$.

5. If $X_1 \sim \mathcal{N}(\mu, \sigma^2)$ in the setting of the CLT, then $S_n \sim \mathcal{N}(n\mu, n\sigma^2)$ for every $n$ (since the sum of independent Gaussians is Gaussian, with mean the sum of the means and variance the sum of the variances). So

$$\frac{S_n - n\mu}{\sqrt{n\sigma^2}} \sim \mathcal{N}(0,1)$$

for every $n$. In particular it converges in distribution to a standard normal as $n \to \infty$!

# 3 Conditional Expectation

Complementary reading: [1, §1.4] and [2, Appendix A].

## 3.1 Preliminary definitions

**Definition 3.1** (Partition)**.** *Suppose that* $(\Omega, \mathcal{F}, \mathbb{P})$ *is a probability space and* $B_1, B_2, \ldots \in \mathcal{F}$*. The collection* $(B_i)_{i\geq 1}$ *is said to form a* ***partition*** *of* $\Omega$ *if:*

- *they are* ***pairwise disjoint****, i.e.,* $B_i \cap B_j = \emptyset$ *if* $i \neq j$*;*
- *they cover* $\Omega$*, i.e.,* $\cup_k B_k = \Omega$*.*

**Lemma 3.2** (Law of total probability)**.** *If* $(B_i)_{i\geq 1}$ *form a partition of* $(\Omega, \mathcal{F}, \mathbb{P})$ *and* $A \in \mathcal{F}$*, then*

$$\mathbb{P}(A) = \sum_{i\geq 1} \mathbb{P}(A|B_i)\mathbb{P}(B_i). \tag{61}$$

*Proof.* Since the $(B_i)_i$ cover $\Omega$ it holds that $A = \cup_{i\geq 1}(A \cap B_i)$, and since they form a partition, these are disjoint events. It therefore follows that $\mathbb{P}(A) = \sum_{i\geq 1} \mathbb{P}(A \cap B_i) = \mathbb{P}(A|B_i)\mathbb{P}(B_i)$, where the second equality comes from the definition of conditional probability. □

**Definition 3.3** (Expectation given an event)**.** *Suppose that* $X$ *is an integrable random variable on a probability space* $(\Omega, \mathcal{F}, \mathbb{P})$*, and* $A \in \mathcal{F}$*. Then the conditional expectation of* $X$ *given* $A$ *is defined by*

$$\mathbb{E}(X|A) = \frac{\mathbb{E}(X\mathbf{1}_A)}{\mathbb{P}(A)}. \tag{62}$$

Note that this coincides with the definition of conditional probability when $X = \mathbf{1}_B$ for an event $B$.

**Example 3.4.** *Suppose that* $X$ *is a discrete random variable on* $(\Omega, \mathcal{F}, \mathbb{P})$ *taking values in* $\mathcal{X} \subset \mathbb{R}$*. Then for* $A \in \mathcal{F}$*:*

$$\mathbb{E}(X|A) = \sum_{x\in\mathcal{X}} x\mathbb{P}(X = x|A). \tag{63}$$

*Compare this expression with the usual definition of expectation for discrete random variables.*

**Lemma 3.5** (Law of total expectation)**.** *If* $(B_i)_{i\geq 1}$ *form a partition of* $(\Omega, \mathcal{F}, \mathbb{P})$ *and* $X$ *is an integrable random variable defined on* $(\Omega, \mathcal{F}, \mathbb{P})$ *then*

$$\mathbb{E}(X) = \sum_{i\geq 1} \mathbb{E}(X|B_i)\mathbb{P}(B_i). \tag{64}$$

*Proof.* When $X$ is a positive discrete random variable, this follows from the law of total probability. For general $X$ this follows from the definition of expectation (after a bit of work). □

### Exercises from §3.1

**Exercise 3.1.** *Suppose you have two fair (6-sided) dice, one red and the other blue. The two dice are rolled independently. Let* $X$ *be the score on the red die and* $Y$ *be the score on the blue die. Calculate the following.*

*(i)* $\mathbb{E}(X \mid X$ *is even*$)$*;*

*(ii)* $\mathbb{E}(X \mid X$ *is odd*$)$*;*

*(iii)* $\mathbb{E}(X + Y \mid X + Y$ *is even*$)$*;*

*(iv)* $\mathbb{E}(X + Y \mid X + Y$ *is odd*$)$*.*

### 3.2 Conditioning on a $\sigma$-field

The next step is to define conditional expectation of a random variable given a $\sigma$-field, rather than given an event. Informally, this is the expected value of the random variable given all the information contained in the $\sigma$-field. In particular, it is a **random variable** (that is measurable with respect to the $\sigma$-field in question) and not just a number (like ordinary expectation).

To begin, suppose that $(B_i)_{i\geq 1}$ form a countable partition of $(\Omega, \mathcal{F}, \mathbb{P})$. Let $\mathcal{G}$ be the $\sigma$-field generated by $(B_i)_{i\geq 1}$. If $X$ is an integrable random variable on $(\Omega, \mathcal{F}, \mathbb{P})$ then the **conditional expectation of** $X$ **given** $\mathcal{G}$ is defined by:

**Definition 3.6** (Conditional expectation given a partition)**.**

$$\mathbb{E}(X|\mathcal{G}) = \sum_{i=1} \mathbb{E}(X|B_i)\mathbf{1}_{B_i}, \tag{65}$$

*where the terms* $\mathbb{E}(X|B_i)$ *are defined by Definition 3.3. So this is a random variable that is equal to* $\mathbb{E}(X|B_i)$ *on the event* $B_i$ *for every* $i$*.*

**Example 3.7.** *Suppose that* $\Omega = [0,1]$*,* $\mathcal{F} = \mathfrak{B}([0,1])$ *and* $\mathbb{P}((a,b)) = b-a$ *for all* $(a,b) \subset [0,1]$*. Let* $U(\omega) = \omega$ *so that* $U$ *is a uniform random variable on* $[0,1]$ *(like we saw in §1). Then if* $B_i = [\frac{i-1}{n}, \frac{i}{n}]$ *for* $1 \leq i \leq n$*,*

$$\mathbb{E}(U|B_i) = \frac{\mathbb{E}(U\mathbf{1}_{B_i})}{\mathbb{P}(B_i)} = n. \int_{(i-1)/n}^{i/n} dx = (i - \frac{1}{2}) \tag{66}$$

*for each* $i$ *(using in the second equality that the probability density function for* $U$ *is just* $1$ *on* $[0,1]$ *and* $0$ *elsewhere). So if* $\mathcal{G} = \sigma(\{B_1, \cdots, B_n\})$ *then*

$$\mathbb{E}(U|\mathcal{G}) = \sum_{i=1}^{n} (i - \frac{1}{2})\mathbf{1}_{B_i}. \tag{67}$$

*In other words,* $\mathbb{E}(U|\mathcal{G}) : [0,1] \to \mathbb{R}$ *is the random variable that is equal to the "middle" of the interval* $B_i$ *on each* $B_i$*.*

The general definition is a bit more abstract.

**Definition 3.8** (Conditional expectation given a $\sigma$-field)**.** *Suppose that* $(\Omega, \mathcal{F}, \mathbb{P})$ *is a probability space,* $\mathcal{G} \subset \mathcal{F}$ *is a sub* $\sigma$*-field, and* $X$ *is an integrable random variable. Then there exists a unique integrable random variable* $Y$ *such that:*

- $Y$ *is* $\mathcal{G}$*-measurable;*
- $\mathbb{E}(Y\mathbf{1}_A) = \mathbb{E}(X\mathbf{1}_A)$ *for every* $A \in \mathcal{G}$*.*

*This unique random variable* $Y$ *is called the* ***conditional expectation of*** $X$ ***given*** $\mathcal{G}$ *and is denoted by*

$$\mathbb{E}(X|\mathcal{G}). \tag{68}$$

The existence of this random variable can be deduced via approximation from the case where $\mathcal{G} = \sigma((B_i)_{i\geq 1})$ for $(B_i)_i$ a partition, but this will not be done here. See [1, Appendix 6] for a proof.

**Example 3.9.** *Consider again the example where* $X$ *is the outcome of a fair die roll and* $\mathcal{G} = \sigma(E)$ *where* $E$ *is the event that the roll is even. Then* $\mathcal{G}$ *is generated by the partition* $E, E^c$ *of* $\Omega$ *so*

$$\mathbb{E}(X|\mathcal{G}) = \mathbb{E}(X|E)\mathbf{1}_E + \mathbb{E}(X|E^c)\mathbf{1}_{E^c} = \frac{2+4+6}{3}\mathbf{1}_E = \frac{1+3+5}{3}\mathbf{1}_{E^c} = 4\mathbf{1}_E + 3\mathbf{1}_{E^c}. \tag{69}$$

*That is,* $\mathbb{E}(X|\mathcal{G})$ *is the random variable equal to* $4$ *on the event that the dice roll is even and equal to* $3$ *on the event that it is odd.*

**Conditioning on a random variable**

Finally, let us define **conditional expectation given a random variable**, which is a special case of conditional expectation given a $\sigma$-algebra.

**Definition 3.10.** *Suppose that $X$ is an integrable random variable on $(\Omega, \mathcal{F}, \mathbb{P})$ and that $Y$ is another random variable on $(\Omega, \mathcal{F}, \mathbb{P})$. Then the **conditional expectation of** $X$ **given** $Y$ is defined by*

$$\mathbb{E}(X|Y) = \mathbb{E}(X|\mathcal{G})\ ;\ \mathcal{G} = \sigma(Y). \tag{70}$$

**Example 3.11.** *If $X, Y$ are discrete random variables taking values in $\mathcal{X}, \mathcal{Y}$ respectively, then*

$$\mathbb{E}(X|Y) = \sum_{y\in\mathcal{Y}} \mathbf{1}_{\{Y=y\}}\mathbb{E}(X|Y=y) = \sum_{y\in\mathcal{Y}} \left( \sum_{x\in\mathcal{X}} x\mathbb{P}(X=x|Y=y) \right) \tag{71}$$

### Exercises from §3.2

**Exercise 3.2.** *In the set up of Definition 3.6, verify that:*

- $\mathbb{E}(X|\mathcal{G})$ *is $\mathcal{G}$-measurable;*
- $\mathbb{E}(X|\mathcal{G})$ *is integrable and* $\mathbb{E}(\mathbb{E}(X|\mathcal{G})) = \mathbb{E}(X)$*;*
- *for any* $A \in \mathcal{G}$ $\mathbb{E}(X\mathbf{1}_A) = \mathbb{E}(\mathbb{E}(X|\mathcal{G})\mathbf{1}_A)$*.*

**Exercise 3.3.** *When $\mathcal{G} = \{\emptyset, \Omega\}$ show that $\mathbb{E}(X|\mathcal{G})$ is just equal to $\mathbb{E}(X)$.*

**Exercise 3.4.** *Take the set-up of Exercise 3.1. What is $\mathbb{E}(X|Y)$? What about $\mathbb{E}(X+Y|Y)$?*

**Exercise 3.5.** *Suppose that a bacteria colony starts with a single bacterium and every second, any bacterium alive in the colony either dies with probability $1-p$ or splits into two with probability $p$. Write $Z_n$ for the size of the colony after $n$ seconds (so $Z_0 = 1$). Let $\mathcal{G}_n$ be the $\sigma$-algebra generated by the first $n$ seconds of activity, i.e., by $Z_0, \ldots, Z_n$. What is $\mathbb{E}(Z_{n+1}|\mathcal{G}_n)$?*

## 3.3 Properties of conditional expectation

**Lemma 3.12** (Basic properties)**.** *Let $X, Y$ be integrable random variables on $(\Omega, \mathcal{F}, \mathbb{P})$ and $\mathcal{G} \subset \mathcal{F}$ a $\sigma$-field. Then:*

*(i)* $\mathbb{E}(\mathbb{E}(X|\mathcal{G})) = \mathbb{E}(X)$*;*

*(ii) if $X$ is $\mathcal{G}$-measurable then $\mathbb{E}(X|\mathcal{G}) = X$ a.s.;*

*(iii) if $X$ is independent of $\mathcal{G}$ then $\mathbb{E}(X|\mathcal{G}) = \mathbb{E}(X)$ a.s.;*

*(iv) if $X \geq 0$ then $\mathbb{E}(X|\mathcal{G}) \geq 0$ a.s.;*

*(v) for any $a, b \in \mathbb{R}$, $\mathbb{E}(aX + bY|\mathcal{G}) = a\mathbb{E}(X|\mathcal{G}) + b\mathbb{E}(Y|\mathcal{G})$;*

**Lemma 3.13** (Conditional Jensen's inequality)**.** *If $X$ is a random variable on $(\Omega, \mathcal{F}, \mathbb{P})$ and $\phi : \mathbb{R} \to \mathbb{R}$ is a convex function such that $\phi(X)$ is integrable, $\mathcal{G} \subset \mathcal{F}$ a sub $\sigma$-field, then*

$$\phi(\mathbb{E}(X|\mathcal{G})) \leq \mathbb{E}(\phi(X)|\mathcal{G}) \text{ a.s.} \tag{72}$$

Observe that taking $\mathcal{G} = \{\emptyset, \Omega\}$ gives the unconditional version of Jensen's inequality. If $\phi$ is convex and $\phi(X)$ is integrable then

$$\phi(\mathbb{E}(X)) \leq \mathbb{E}(\phi(X)). \tag{73}$$

*Proof.* Let us admit without proof that any convex function $\phi$ is the supremum of countably many affine functions. That is, $\phi(x) = \sup_k(a_k x + b_k)$ for all $x$, for some countable collection $(a_k, b_k)_k$. In particular, for each $k$ it holds that $\phi(x) \geq a_k x + b_k$. By Lemma 3.12(iv) and (v), it therefore follows that $\mathbb{E}(\phi(X)|\mathcal{G}) \geq a_k \mathbb{E}(X|\mathcal{G}) + b_k$ a.s. for each $k$, and so (since the supremum is over a countable set)

$$\mathbb{E}(\phi(X)|\mathcal{G}) \geq \sup_k (a_k \mathbb{E}(X|\mathcal{G}) + b_k) = \phi(\mathbb{E}(X|\mathcal{G})) \text{ a.s.} \tag{74}$$

as desired. □

Note in particular this implies

$$|\mathbb{E}(X|\mathcal{G})|^p \leq \mathbb{E}(|X|^p|\mathcal{G}) \text{ a.s.} \tag{75}$$

for any $p \geq 1$, and so by Lemma 3.12(i) if $Y = \mathbb{E}(X|\mathcal{G})$ then

$$\mathbb{E}(|Y|^p) \leq \mathbb{E}(|X|^p). \tag{76}$$

This property is sometimes described by saying that **conditioning is a contraction** for $p$th moments.

Let us also mention here that the results concerning convergence of expectations (Fatou's lemma, the monotone convergence theorem, the dominated convergence theorem) have conditional versions (that are exactly the same, replacing expectations by their conditional counterparts).

**Lemma 3.14** (Tower law)**.** *Suppose that* $(\Omega, \mathcal{F}, \mathbb{P})$ *is a probability space,* $\mathcal{H} \subset \mathcal{G}$ *are sub* $\sigma$*-fields and* $X$ *is an integrable random variable. Then*

$$\mathbb{E}(\mathbb{E}(X|\mathcal{G})|\mathcal{H}) = \mathbb{E}(X|\mathcal{H}) \text{ a.s.} \tag{77}$$

*Proof.* By definition of conditional expectation $\mathbb{E}(\mathbb{E}(X|\mathcal{G})|\mathcal{H})$, it suffices to check that $\mathbb{E}(X|\mathcal{H})$ is $\mathcal{H}$-measurable and $\mathbb{E}(\mathbb{E}(X|\mathcal{G})\mathbf{1}_A) = \mathbb{E}(\mathbb{E}(X|\mathcal{H})\mathbf{1}_A)$ for all $A \in \mathcal{H}$. The first property holds by definition of $\mathbb{E}(X|\mathcal{H})$. The second holds since any such $A$ is both $\mathcal{H}$- and $\mathcal{G}$-measurable, meaning that $\mathbb{E}(\mathbb{E}(X|\mathcal{H})\mathbf{1}_A) = \mathbb{E}(X\mathbf{1}_A)$ by definition of $\mathbb{E}(X|\mathcal{H})$ and $\mathbb{E}(\mathbb{E}(X|\mathcal{G})\mathbf{1}_A) = \mathbb{E}(X\mathbf{1}_A)$ by definition of $\mathbb{E}(X|\mathcal{G})$. □

**Lemma 3.15** (Taking out what is known)**.** *Suppose that* $(\Omega, \mathcal{F}, \mathbb{P})$ *is a probability space,* $\mathcal{G}$ *is a sub* $\sigma$*-field,* $X$ *is an integrable random variable, and* $Y$ *is a bounded and* $\mathcal{G}$*-measurable random variable. Then*

$$\mathbb{E}(XY|\mathcal{G}) = Y\mathbb{E}(X|\mathcal{G}) \text{ a.s.} \tag{78}$$

*Proof.* First suppose that $Y = \mathbf{1}_B$ for some $B \in \mathcal{G}$. Then it is immediate that $Y\mathbb{E}(X|\mathcal{G})$ is $\mathcal{G}$-measurable, and for any $A \in \mathcal{G}$ it holds that $\mathbb{E}(\mathbf{1}_B\mathbb{E}(X|\mathcal{G})\mathbf{1}_A) = \mathbb{E}(\mathbf{1}_{A\cap B}\mathbb{E}(X|\mathcal{G})) = \mathbb{E}(X\mathbf{1}_{A\cap B}) = \mathbb{E}(XY\mathbf{1}_A)$. Thus the lemma is proved in this case. The result extends to positive random variables $X, Y$ by linearity of expectation and the monotone convergence theorem (approximating $X, Y$ by finite weighted sums of indicator functions). The general case follows by writing $X = X^+ - X^-$ and $Y = Y^+ - Y^-$ with $X^\pm, Y^\pm$ non-negative. □

**Remark 3.16.** The assumption that $Y$ is bounded, is to ensure that $XY$ is integrable. In fact, by the Cauchy–Schwarz inequality, it is enough to assume that instead $\mathbb{E}(X^2), \mathbb{E}(Y^2) < \infty$. (There is also an extension to the Cauchy–Schwarz inequality called Hölder's inequality, which implies that $\mathbb{E}(X^p) < \infty, \mathbb{E}(Y^q) < \infty$ for any $(1/p) + (1/q) = 1$ is enough).

## Exercises from §3.3

**Exercise 3.6.** *Prove the properties in Lemma 3.12.*

**Exercise 3.7.** *Going back to the setting of Exercise 3.5 from §3.2, what is* $\mathbb{E}(Z_{n+m}|\mathcal{G}_n)$ *for* $m > 1$*?*

**Solutions to exercises from §3**

1. (i) For $i = 1, 3, 5$, $\mathbb{P}(X = i|X \text{ is even}) = \mathbb{P}(X = i \text{ and } X \text{ is even})/\mathbb{P}(X \text{ is even}) = 0/(1/2) = 0$ since $X$ cannot be equal to $1, 3$ or $5$ and also be even. On the other hand, for $i = 2, 4, 6$, $\mathbb{P}(X = i|X \text{ is even}) = \mathbb{P}(X = i \text{ and } X \text{ is even})/\mathbb{P}(X \text{ is even}) = (1/6)/(1/2) = 1/3$, since $\mathbb{P}(X = i \text{ and } X \text{ is even})$ is just $\mathbb{P}(X = i) = 1/6$ for $i = 2, 4, 6$. Therefore, by definition

$$\mathbb{E}(X|X \text{ is even}) = 1.0 + 2.(1/3) + 3.0 + 4.(1/3) + 5.0 + 6.(1/3) = \frac{2+4+6}{3} = 4.$$

(ii) Similarly

$$\mathbb{E}(X|X \text{ is odd}) = 1.(1/3) + 2.0 + 3.(1/3) + 4.0 + 5.(1/3) + 6.0 = \frac{1+3+5}{3} = 3.$$

(iii) Now $\mathbb{P}(X + Y = 2) = \mathbb{P}(X + Y = 12) = 1/36$, $\mathbb{P}(X + Y = 3) = \mathbb{P}(X + Y = 11) = 2/36$, $\mathbb{P}(X+Y = 4) = \mathbb{P}(X+Y = 10) = 3/36$, $\mathbb{P}(X+Y = 5) = \mathbb{P}(X+Y = 9) = 4/36$, $\mathbb{P}(X+Y = 6) = \mathbb{P}(X + Y = 8) = 5/36$ and $\mathbb{P}(X + Y = 7) = 6/36$. Therefore, $\mathbb{P}(X + Y \text{is even}) = 18/36 = 1/2$ and

$$\mathbb{E}(X + Y|X + Y \text{ is even}) = 2.\frac{1/36}{1/2} + 4.\frac{3/36}{1/2} + 6.\frac{5/36}{1/2} + 8.\frac{5/36}{1/2} + 10.\frac{3/36}{1/2} + 12.\frac{1/36}{1/2} = 7.$$

(iv) Similarly,

$$\mathbb{E}(X + Y|X + Y \text{ is odd}) = 3.\frac{2/36}{1/2} + 5.\frac{4/36}{1/2} + 7.\frac{6/36}{1/2} + 9.\frac{4/36}{1/2} + 11.\frac{2/36}{1/2} = 7.$$

Note that by symmetry $\mathbb{E}(X + Y) = 7$ and so we didn't really need to do (iv) once we knew the answer to (iii), because by the law of total expectation

$$\begin{aligned} 7 = \mathbb{E}(X + Y) &= \mathbb{E}(X + Y|X + Y \text{ is even})\mathbb{P}(X + Y \text{ is even}) + \mathbb{E}(X + Y|X + Y \text{ is odd})\mathbb{P}(X + Y \text{ is odd}) \\ &= 7\frac{1}{2} + \mathbb{E}(X + Y|X + Y \text{ is odd})\frac{1}{2} \end{aligned}$$

and we could have already deduced that $\mathbb{E}(X + Y|X + Y \text{ is odd}) = 7$.

2. We need to check that $Y := \sum_{i\geq 1} \mathbb{E}(X|B_i)\mathbf{1}_{B_i}$ satisfies the three defining properties of conditional expectation with respect to $\mathcal{G} = \sigma(\{B_i\}_{i\geq 1})$. First, $Y$ is clearly $\mathcal{G}$-measurable since it is a weighted sum of indicator functions of events in $\mathcal{G}$. It is also integrable since $|Y| = |\sum_{i\geq 1} \mathbb{E}(X|B_i)\mathbf{1}_{B_i}| \leq \sum_{i\geq 1} \mathbb{E}(|X||B_i)\mathbf{1}_{B_i}$ and the right hand side is the increasing limit of $Z_n : \sum_{i=1}^n \mathbb{E}(|X||B_i)\mathbf{1}_{B_i}$, so by the monotone convergence theorem has finite expectation equal to $\lim_n \mathbb{E}(Z_n) = \lim_n \sum_{i=1}^n \mathbb{E}(|X||B_i)\mathbb{P}(B_i) = \lim_n \mathbb{E}(|X|\mathbf{1}_{B_1\cdots\cup\ldots B_n}) = \mathbb{E}(|X|)$. Similarly we have

$$\mathbb{E}(Y) = \mathbb{E}(\sum_{i\geq 1} \mathbb{E}(X|B_i)\mathbf{1}_{B_i}) = \sum_{i\geq 1} \mathbb{E}(\mathbb{E}(X|B_i)\mathbf{1}_{B_i}) = \sum_{i\geq 1} \mathbb{E}(X|B_i)\mathbb{P}(B_i) = \mathbb{E}(X)$$

by the law of total expectation. Finally, we need to check that $\mathbb{E}(Y\mathbf{1}_A) = \mathbb{E}(X\mathbf{1}_A)$ for any $A \in \mathcal{G}$. However, any $A \in \mathcal{G}$ is a union of events $B_i$ for some collection of $i$; therefore it suffices to check the claim with $A = B_j$ for some fixed $j$. But then

$$\mathbb{E}(Y\mathbf{1}_{B_j}) = \mathbb{E}(\sum_{i\geq 1} \mathbb{E}(X|B_i)\mathbf{1}_{B_i}\mathbf{1}_{B_j}) = \mathbb{E}(\mathbb{E}(X|B_j)\mathbf{1}_{B_j}) = \mathbb{E}(X|B_j)\mathbb{P}(B_j) = \mathbb{E}(X\mathbf{1}_{B_j})$$

as required, where the first equality follows since $\mathbf{1}_{B_j}\mathbf{1}_{B_i} = \mathbf{1}_{B_j}$ if $j = i$ and 0 if $i \neq j$.

3. We need to check that the constant variable $Y := \mathbb{E}(X)$ satisfies the three defining properties of conditional expectation with respect to $\mathcal{G} = \{\emptyset, \Omega\}$. First, $Y$ is measurable with respect to $\mathcal{G}$ since $Y^{-1}((-\infty, x]) = \{\omega \in \Omega : Y(w) \leq x\} = \{\omega : \mathbb{E}(X) \leq x\}$ which is the empty set if $x < \mathbb{E}(X)$ and $\Omega$ otherwise. Secondly it is integrable, since it is constant, with $\mathbb{E}(Y) = \mathbb{E}(X)$. Finally, if $A \in \mathcal{G}$ then $A = \emptyset$ or $A = \Omega$, so we just need to check that $\mathbb{E}(Y\mathbf{1}_\emptyset) = \mathbb{E}(X\mathbf{1}_\emptyset)$ and $E(Y\mathbf{1}_\Omega) = \mathbb{E}(X\mathbf{1}_\Omega)$. This is immediate: the first equality follows since $\mathbf{1}_\emptyset \equiv 0$ and the second since $\mathbf{1}_\Omega \equiv 1$. Note that this should be interpreted as saying that $\mathbb{E}(X|\text{"no information"}) = \mathbb{E}(X)$.

4. Since $X$ and $Y$ are independent, we have $\mathbb{E}(X|Y) = \mathbb{E}(X)$. More concretely, $\mathbb{E}(X|Y) = \sum_{i=1}^{6} \mathbb{E}(X|Y = i)\mathbf{1}_{Y=i}$ where $\mathbb{E}(X|Y = i) = \sum_{j=1}^{6} j\mathbb{P}(X = j|Y = i) = \sum_{j=1}^{6} j/6 = 3.5$ for each $i$ (by independence) $\Rightarrow \mathbb{E}(X|Y) = \sum_{i=1}^{6} 3.5\mathbf{1}_{Y=i} \equiv 3.5 = \mathbb{E}(X)$.

   It is straightforward to check from the abstract definition that $\mathbb{E}(Z_1 + Z_2|\mathcal{G}) = \mathbb{E}(Z_1|\mathcal{G}) + \mathbb{E}(Z_2|\mathcal{G})$ for any integrable $Z_1, Z_2$. Therefore $\mathbb{E}(X + Y|Y) = \mathbb{E}(X|Y) + \mathbb{E}(Y|Y) = 3.5 + Y$. (One can also check that $3.5 + Y$ does satisfy the defining properties of conditional expectation, and therefore must be the right answer).

5. For each bacteria in the colony at time $n$, there will be an average of $2.p + 0.(1 - p) = 2p$ bacteria in the colony at time $n + 1$. It is therefore natural to guess that $\mathbb{E}(Z_{n+1}|\mathcal{G}_n) = 2pZ_n$. So let us check that $2pZ_n$ satisfies the properties of conditional expectation. Firstly, it is $\mathcal{G}_n$ measurable, since it is just a multiple of $Z_n$. As $Z_n \leq 2^n Z_0$, it is also bounded and therefore integrable, with expectation

$$\begin{aligned}
2p\mathbb{E}(Z_n) &= 2p\mathbb{E}(\sum_{i\geq 1} \mathbf{1}_{i\leq Z_n}) \\
&= 2\sum_{i\geq 1} \mathbb{P}(i \leq Z_n)\mathbb{P}(\text{"ith bacterium at stage n splits into two"}) \\
&= 2\sum_{i\geq 1} \mathbb{P}(i \leq Z_n \cap \text{"ith bacterium at stage n splits into two"}) \\
&= \mathbb{E}(\sum_{i\geq 1} 2.\mathbf{1}_{\{i\leq Z_n \cap \text{"ith bacterium at stage n splits into two"}\}}) \\
&= \mathbb{E}(Z_{n+1})
\end{aligned}$$

   where the third line follows by independence. We can similarly check that $\mathbb{E}(2pZ_n\mathbf{1}_A) = \mathbb{E}(Z_{n+1}\mathbf{1}_A)$ for any $A \in \mathcal{G}_n$, and hence $\mathbb{E}(Z_{n+1}|\mathcal{G}_n)$ is indeed equal to $2pZ_n$.

6. (i) This follows from the second defining property of conditional expectation, taking $A = \Omega$.

   (ii) If $X$ is $\mathcal{G}$ measurable, then $Y = X$ is clearly integrable, $\mathcal{G}$-measurable, and satsifies $\mathbb{E}(X\mathbf{1}_A) = \mathbb{E}(Y\mathbf{1}_A)$ for any $A \in \mathcal{G}$. Therefore $\mathbb{E}(X|\mathcal{G})$ must be equal to $X = Y$.

   (iii) $\mathbb{E}(X)$ is a $\mathcal{G}$-measurable random variable for any $\mathcal{G}$ since $\{\omega : \mathbb{E}(X)(w) \leq x\}$ is either $\Omega$ (if $x \geq \mathbb{E}(X)$) or $\emptyset$ (otherwise), and $\emptyset, \Omega$ must be elements of the $\sigma$-field $\mathcal{G}$. It is also clearly integrable, and for any $A$ which is $\mathcal{G}$-measurable $\mathbb{E}(\mathbb{E}(X)\mathbf{1}_A) = \mathbb{E}(X)\mathbb{P}(A)$. By independence of $X$ and $\mathbf{1}_A$, this last expression is equal to $\mathbb{E}(X\mathbf{1}_A)$, so indeed $\mathbb{E}(X|\mathcal{G}) = \mathbb{E}(X)$.

   (iv) If $X \geq 0$, set $\mathbb{E}(X|\mathcal{G}) = Y$ so that $Y$ is $\mathcal{G}$-measurable by definition. Then the event $A = \{Y < 0\}$ must be an element of $\mathcal{G}$, and therefore $\mathbb{E}(X\mathbf{1}_A) = \mathbb{E}(Y\mathbf{1}_A)$. Now the right-hand side is clearly negative, so the left-hand side must be negative. But the left-hand side is also positive since $X\mathbf{1}_A \geq 0$ a.s. Therefore we must have $\mathbb{E}(Y\mathbf{1}_{\{Y<0\}}) = 0$. This implies that $\mathbb{E}(X|\mathcal{G}) = Y \geq 0$ a.s.

   (v) We need to check that $Z := a\mathbb{E}(X|\mathcal{G}) + b\mathbb{E}(Y|\mathcal{G})$ satisfies the defining properties for the conditional expectation $\mathbb{E}(aX + bY|\mathcal{G})$. Since $\mathbb{E}(X|\mathcal{G})$ and $\mathbb{E}(Y|\mathcal{G})$ are $\mathcal{G}$-measurable by definition, $Z$ is also $\mathcal{G}$-measurable. Similarly it is integrable as a sum of integrable random variables. Finally if $A \in \mathcal{G}$, then $\mathbb{E}(Z\mathbf{1}_A) = \mathbb{E}(a\mathbb{E}(X|\mathcal{G})\mathbf{1}_A + b\mathbb{E}(Y|\mathcal{G})\mathbf{1}_A) = a\mathbb{E}(\mathbb{E}(X|\mathcal{G})\mathbf{1}_A) + b\mathbb{E}(\mathbb{E}(Y|\mathcal{G})\mathbf{1}_A)$ just by linearity of (non-conditional) expectation. But this is equal to $a\mathbb{E}(X\mathbf{1}_A) + b\mathbb{E}(Y\mathbf{1}_A)$ by definition of $\mathbb{E}(X|\mathcal{G})$ and $\mathbb{E}(Y|\mathcal{G})$, and finally by linearity again, this is equal to $\mathbb{E}((aX+bY)\mathbf{1}_A)$. This completes the check.

7. Recall that we showed in a previous exercise that $\mathbb{E}(Z_{n+1}|\mathcal{G}_n) = 2pZ_n$ for any $n$. We will use the tower law to repeat this step and deduce that $\mathbb{E}(Z_{n+m}|\mathcal{G}_n) = (2p)^m Z_n$ for any $n$ and $m \geq 2$. More precisely, since $\mathcal{G}_n \subset \mathcal{G}_{n+m-1}$ for $m \geq 2$, we can write

$$\mathbb{E}(Z_{n+m}|\mathcal{G}_n) = \mathbb{E}(\mathbb{E}(Z_{n+m}|\mathcal{G}_{n+m-1})|\mathcal{G}_n) = 2p\mathbb{E}(Z_{n+m-1}|\mathcal{G}_n)$$

by the tower law and the aforementioned "one-step" result. Repeating this $m-1$ times we obtain that

$$\mathbb{E}(Z_{n+m}|\mathcal{G}_n) = (2p)^{m-1}\mathbb{E}(Z_{n+1}|\mathcal{G}_n) = (2p)^m Z_n.$$

# 4 Stochastic processes

Complementary reading: [1, §1.2] and [2, Appendices B & C].

## 4.1 Stochastic processes in discrete and continuous time

**Definition 4.1** (Stochastic processes in time). *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $\mathcal{T}$ be a subset of $\mathbb{R}$. Then a* ***stochastic process***

$$\mathfrak{X} = (X_t)_{t\in\mathcal{T}} \tag{79}$$

*is a collection of random variables $X_t$ for each $t \in \mathcal{T}$. That is, for any $t \in \mathcal{T}$,*

$$X_t : \Omega \to \mathbb{R} \tag{80}$$

*is a random variable (= $\mathcal{F}$-measurable function).*

So for each fixed $\omega \in \Omega$, the map $X(\omega)$ taking $t \mapsto X_t(\omega)$ is a function from $\mathcal{T} \to \mathbb{R}$, or an element of $\mathbb{R}^{\mathcal{T}}$. This is called a **trajectory** or **sample path** of the stochastic process.

**Caution/Aside**: one can ask if the map taking $\omega$ to the associated trajectory $X$ or function above is **measurable** in some sense. That is, if some topology is put on the space of functions $\mathbb{R}^{\mathcal{T}}$ from $\mathcal{T} \to \mathbb{R}$, does it hold that $\{\omega \in \Omega : X(\omega) \in A\} \in \mathcal{F}$ for all open sets $A$ of $\mathbb{R}^{\mathcal{T}}$. This, naturally, is not the case for an arbitrary topology on $\mathbb{R}^{\mathcal{T}}$. It is however the case when $\mathbb{R}^{\mathcal{T}}$ is given the **product topology**; that is the topology generated by sets of the form $\{(Y(t))_{t\in\mathcal{T}} \in \mathbb{R}^{\mathcal{T}} : Y(t_1) \in (a_1, b_1), \ldots, Y(t_n) \in (a_n, b_n)\}$ for some $n \in \mathbb{N}$, $t_1, \ldots, t_n \in \mathcal{T}$ and $a_1, b_1, \ldots, a_n, b_n \in \mathbb{R}$. This is what makes the **finite dimensional distributions** of a stochastic process (to be discussed below) particularly important.

**Definition 4.2** (Discrete and continuous time). *In this course, if $\mathcal{T} = \{0, 1, 2, 3, \ldots\}$ then a stochastic process $\mathfrak{X} = (X_t)_{t\in\mathcal{T}} = (X_1, X_2, \ldots)$ will be referred to as a* ***discrete time stochastic process****. In the general case, usually this will be $\mathcal{T} = [0, \infty)$ or $\mathcal{T} = [0, 1]$, $\mathfrak{X}$ will be referred to as a* ***continuous time stochastic process****.*

**Example 4.3.** *A classic example of a discrete time stochastic process is the* ***simple symmetric random walk****. This is defined from an i.i.d. collection $X_1, X_2, \ldots$ of random variables with $\mathbb{P}(X_i = 1) = \mathbb{P}(X_i = -1) = 1/2$ for $i = 1, 2, \ldots$. The random walk at time $n$ is then $S_n = \sum_{i=1}^{n} X_i$. So this is the discrete time process ("walker") that moves left or right at every time step with equal probability.*

*A continuous time version of this process can be defined by letting the moves $X_i$ occur not at fixed integer times, but after independent exponential waiting times. More precisely, along with $X_1, X_2, \ldots$, take a collection $T_1, T_2, \ldots$ of i.i.d.* Exp(1) *random variables. Then for $t \in [0, \infty)$ setting $N_t = \max\{n : T_1 + \ldots + T_n < t\}$, one can define the continuous time random walk $S_t = \sum_{i=1}^{N_t} X_i$.*

**Definition 4.4** (Finite dimensional distributions). *The* ***finite dimensional distributions (fidis)*** *of a stochastic process $\mathfrak{X} = (X_t)_{t\in\mathcal{T}}$ are the distributions of the random vectors*

$$(X_{t_1}, \ldots, X_{t_n}) \tag{81}$$

*for $n \in \mathbb{N}$ and $(t_1, \ldots, t_n) \in \mathbb{R}^n$.*

Two stochastic processes $\mathfrak{X} = (X_t)_{t\in\mathcal{T}}$ and $\mathfrak{X}' = (X'_t)_{t\in\mathcal{T}}$ are said to be **versions** of one another if for every $t \in \mathcal{T}$, $\mathbb{P}(X_t = X'_t) = 1$. Two such versions have the **same** finite dimensional distributions.

**Caution/Aside:** it is possible to define two stochastic processes on the same probability space that are versions of one another (and have the same finite dimensional distributions) but do not have the same sample path properties. For example, let a uniform random variable $U$ on $[0, 1]$ be defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Define $X_t = 0$ and $X'_t = \mathbf{1}_{\{U=t\}}$ for $t \in [0, 1]$. Then $\mathfrak{X} = (X_t)_{t\in[0,1]}$ and

$\mathfrak{X}' = (X'_t)_{t\in[0,1]}$ both define stochastic processes that are versions of one another with the same finite dimensional distributions, but the sample paths of $X$ and $X'$ clearly differ. For instance, $X$ defines a continuous function a.s. while $X'$ a.s. does not.

So, if one wants to work with a stochastic properties having specific sample path properties with probability one, then it is necessary to specify which version of the stochastic process is being worked with.

## Properties of stochastic processes

**Definition 4.5** (Expectation of a stochastic process)**.** *The **expectation function** of a stochastic process $\mathfrak{X} = (X_t)_{t\in\mathcal{T}}$ (such that $X_t$ is integrable for every $t \in \mathcal{T}$) is the function $\mu_{\mathfrak{X}} : \mathcal{T} \to \mathbb{R}$ defined by*

$$\mu_{\mathfrak{X}}(t) = \mathbb{E}(X_t) \text{ for } t \in \mathcal{T}. \tag{82}$$

**Definition 4.6** (Covariance of a stochastic process)**.** *The **covariance function** of a stochastic process $\mathfrak{X} = (X_t)_{t\in\mathcal{T}}$ (such that $X_t$ is square-integrable for every $t \in \mathcal{T}$) is the function $c_{\mathfrak{X}} : \mathcal{T} \times \mathcal{T} \to \mathbb{R}$ defined by*

$$c_{\mathfrak{X}}(t,s) = \operatorname{cov}(X_t, X_s) = \mathbb{E}((X_t - \mu_{\mathfrak{X}}(t))(X_s - \mu_{\mathfrak{X}}(s))) \text{ for } t, s \in \mathcal{T}. \tag{83}$$

**Definition 4.7** (Variance of a stochastic process)**.** *The **variance function** of a stochastic process $\mathfrak{X} = (X_t)_{t\in\mathcal{T}}$ (such that $X_t$ is square-integrable for every $t \in \mathcal{T}$) is the function $\sigma_{\mathfrak{X}} : \mathcal{T} \to \mathbb{R}$ defined by*

$$\sigma_{\mathfrak{X}}(t) = \operatorname{var}(X_t) = \mathbb{E}(X_t^2) - \mathbb{E}(X_t)^2 \text{ for } t \in \mathcal{T}. \tag{84}$$

**Example 4.8.** *Let $\mathfrak{X} = (X_t)_{t\in[0,1]}$ be such that $X_t \sim \mathcal{N}(0,1)$ for every $t \in [0,1]$ and $X_s, X_t$ are independent for $s \neq t$. Then for any $t, s \in [0,1]$:*

$$\mu_{\mathfrak{X}}(t) = 0 \; ; \; c_{\mathfrak{X}}(t,s) = 0 \; ; \; \sigma_{\mathfrak{X}}(t) = 1. \tag{85}$$

Now let us describe some special and relevant properties that stochastic processes may possess.

**Definition 4.9** (Strict stationarity)**.** *A stochastic process $\mathfrak{X} = (X_t)_{t\in\mathcal{T}}$ is said to be **strictly stationary** if its finite dimensional distributions do not change under time shifts. That is, if for any $t_1, \ldots, t_n \in \mathcal{T}$ and $h \in \mathbb{R}$ such that $t_1 + h, \ldots, t_n + h \in \mathcal{T}$, the random vectors*

$$(X_{t_1}, \ldots, X_{t_n}) \text{ and } (X_{t_1+h}, \ldots, X_{t_n+h}) \tag{86}$$

*have the same distribution.*

**Definition 4.10** (Stationarity)**.** *A stochastic process $\mathfrak{X} = (X_t)_{t\in\mathcal{T}}$ is said to be **stationary** if for any $t, s \in \mathcal{T}$:*

- *$\mu_{\mathfrak{X}}(t) = \mu_{\mathfrak{X}}(s)$ is equal to some constant; and*
- *$c_{\mathfrak{X}}(t,s)$ depends only on $|s-t|$.*

**Definition 4.11** (Stationarity of increments)**.** *A stochastic process $\mathfrak{X} = (X_t)_{t\in\mathcal{T}}$ is said to have **stationary increments** if for any $t, s \in \mathcal{T}$ and $h \in \mathbb{R}$ such that $t+h, s+h \in \mathcal{T}$,*

$$X_t - X_s \text{ and } X_{t+h} - X_{s+h} \tag{87}$$

*have the same distribution.*

**Definition 4.12** (Independence of increments)**.** *A stochastic process $\mathfrak{X} = (X_t)_{t\in\mathcal{T}}$ **independent increments** if for any $n \in \mathbb{N}$ and $t_1, \ldots, t_n \in \mathcal{T}$,*

$$X_{t_2} - X_{t_1}, \ldots, X_{t_n} - X_{t_{n-1}} \tag{88}$$

*are independent random variables.*

**Example 4.13.** *For example, the random walks from Example 4.3 are have stationary and independent increments. For the discrete time version this follows directly from the definition. For the continuous time version, one needs to use the memorylessness property of the exponential distribution: if* $t, s \geq 0$ *and* $T \sim \mathrm{Exp}(1)$ *then* $\mathbb{P}(T \geq t + s | T \geq t) = \mathbb{P}(T \geq s)$.

*On the other hand, they are not strictly stationary or stationary (for example, the variance is increasing in time).*

### Exercises from §4.1

**Exercise 4.1.** *Compute the mean, variance and covariance functions for the random walks in Example 4.3.*

**Exercise 4.2.** *Consider the process* $(N_t)_{t \in [0,\infty)}$ *from Example 4.3. Compute its mean and covariance functions. Show that the process has independent and stationary increments.*

**Exercise 4.3.** *Show that strict stationarity implies stationarity (in the case of square integrable stochastic processes).*

## 4.2 Types of process

#### Filtrations and adaptedness

**Definition 4.14** (Filtration)**.** *Suppose that* $(\Omega, \mathcal{F}, \mathbb{P})$ *is a probability space, that* $\mathcal{T} \subset \mathbb{R}$ *and that* $\mathcal{F}_t \subset \mathcal{F}$ *is a* $\sigma$*-field for every* $t \in \mathcal{T}$*. The collection* $(\mathcal{F}_t)_{t \in \mathcal{T}}$ *is called a* ***filtration*** *if*

$$\mathcal{F}_s \subset \mathcal{F}_t \text{ whenever } s, t \in \mathcal{T} \text{ are such that } s \leq t. \tag{89}$$

**Example 4.15.** *If* $\mathcal{T} = \{0, 1, 2, \ldots,\}$ *then this is equivalent to the condition that* $\mathcal{F}_t \subset \mathcal{F}_{t+1}$ *for all* $t = 0, 1, 2, \ldots$

If $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space, and $(\mathcal{F}_t)_{t \in \mathcal{T}}$ is a filtration on it, then $(\Omega, \mathcal{F}, (\mathcal{F}_t)_t, \mathbb{P})$ is called a **filtered probability space**.

**Definition 4.16** (Adaptedness)**.** *A stochastic process* $\mathfrak{X} = (X_t)_{t \in \mathcal{T}}$ *is said to be* ***adapted*** *to a a filtration* $(\mathcal{F}_t)_{t \in \mathcal{T}}$ *on* $(\Omega, \mathcal{F}, \mathbb{P})$ *if* $X_t$ *is* $\mathcal{F}_t$*-measurable for each* $t \in \mathcal{T}$.

**Example 4.17** (Natural filtration)**.** *For a given stochastic process* $\mathfrak{X} = (X_t)_{t \in \mathcal{T}}$ *one can define the* $\sigma$*-field* $\mathcal{F}_t$ *generated by* $(X_s)_{s \leq t, s \in \mathcal{T}}$*. That is,* $\mathcal{F}_t := \sigma(\{X_s^{-1}(A); A \in \mathfrak{B}(\mathbb{R}), t \geq s \in \mathcal{T}\})$*.* $\mathfrak{X}$ *will always be adapted to this filtration.*

**Definition 4.18** (Increasing process)**.** *Suppose that* $\mathcal{T} = \{0, 1, 2, ...\}$ *and that* $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \in \mathcal{T}}, \mathbb{P})$ *is a filtered probability space. A stochastic process* $A = (A_t)_{t \in \mathcal{T}}$ *is* ***increasing*** *if:*

- *it is adapted to* $(\mathcal{F}_t)_{t \in \mathcal{T}}$*;*
- $A_t$ *is integrable for all* $t \in \mathcal{T}$*; and*
- $A_0 \leq A_1 \leq \ldots$ *with probability one.*

**Definition 4.19** (Predictable process)**.** *Suppose that* $\mathcal{T} = \{0, 1, 2, ...\}$ *and that* $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \in \mathcal{T}}, \mathbb{P})$ *is a filtered probability space. A stochastic process* $C = (C_t)_{t \in \mathcal{T}}$ *such that* $\sigma(C_t) \subset \mathcal{F}_{t-1}$ *for all* $t \in \mathcal{T}$ *is called a* ***predictable process****.*

Roughly speaking, the value of a predictable process is "known" at the previous time step.

**Example 4.20.** *Any **deterministic** process $(C_t)_{t\in\mathcal{T}}$ (i.e., such that for some $c_0, c_1, \ldots \in \mathbb{R}$, $\mathbb{P}(C_t = c_t \forall\, t \geq 0)$ is predictable. This is because $\sigma(C_t) = \{\emptyset, \Omega\}$ for every $t$, and this is always in $\mathcal{F}_t$ since $\mathcal{F}_t$ is a $\sigma$-field.*

**Definition 4.21** (Continuous and cadlag processes)**.** *Suppose that $\mathfrak{X} = (X_t)_{t\in[0,T]}$ for some $T \in [0, \infty]$. Then $\mathfrak{X}$ is said to be **continuous** if its sample paths or trajectories are continuous functions almost surely.*

*Similarly, $\mathfrak{X}$ is said to be **cadlag** if its sample paths admit left limits and are right-continuous at every point almost surely.*

**Definition 4.22** (Uniform integrability)**.** *A stochastic process $\mathfrak{X} = (X_t)_{t\in\mathcal{T}}$ is said to be **uniformly integrable** if*

$$\sup_{t\in\mathcal{T}} \mathbb{E}(|X_t|\mathbf{1}_{|X_t|\geq K}) \tag{90}$$

*converges to $0$ as $K \to \infty$.*

This will be a useful property later on, which ensures that the process is sufficiently well behaved, and allows one to apply various theorems.

**Example 4.23.** *If $X_t = X$ for all $t \in T$ where $X$ is an integrable random variable, then $\mathfrak{X}$ is uniformly integrable. This is because*

$$\sup_{t\in\mathcal{T}} \mathbb{E}(|X_t|\mathbf{1}_{|X_t|\geq K}) = \mathbb{E}(|X|\mathbf{1}_{|X|\geq K}) = \mathbb{E}(|X|) - \mathbb{E}(|X|\mathbf{1}_{|X|\leq K}) \tag{91}$$

*for each $K$, and the right hand side goes to $0$ as $K \to \infty$ by the monotone convergence theorem.*

### Gaussian processes

In this course, one very important class of stochastic processes are the Gaussian processes. For example, **Brownian motion** (which you will see a great deal of later) is a Gaussian process.

**Definition 4.24** (Gaussian process)**.** *A stochastic process $\mathfrak{X} = (X_t)_{t\in\mathcal{T}}$ is called a **Gaussian process**, if and only if all its finite dimensional distributions are multivariate Gaussian (i.e., the distributions of Gaussian vectors). Equivalently, if and only if for each $n \in \mathbb{N}$, $(a_1, \ldots, a_n) \in \mathbb{R}^n$ and $(t_1, \ldots, t_n) \in \mathcal{T}^n$,*

$$a_1 X_{t_1} + \ldots + a_n X_{t_n} \tag{92}$$

*is a Gaussian random variable.*

**Example 4.25.** *The process $\mathfrak{X}$ from Example 4.8 is a **Gaussian process**. In fact, given $\mathcal{T}$, a function $\mu : \mathcal{T} \to \mathbb{R}$ and a symmetric bilinear form $c : \mathcal{T} \times \mathcal{T} \to \mathbb{R}$ that is non-negative semidefinite (i.e., for every $n \in \mathbb{N}$, $(a_1, \ldots, a_n) \in \mathbb{R}^n$ and $(t_1, \ldots, t_n) \in \mathcal{T}^n$ it holds that $\sum_{i,j=1}^n a_i a_j c(t_i, t_j) \geq 0$), there exists a Gaussian process $\mathfrak{X}$ with $\mu = \mu_{\mathfrak{X}}$ and $c = c_{\mathfrak{X}}$.*

## Exercises from §4.2

**Exercise 4.4.** *If $\mathcal{F}_t = \{\emptyset, \Omega\}$ for every $t \in \mathcal{T}$, show that this defines a filtration. Which processes are measurable with respect to this filtration? What about if $\mathcal{F}_t = \mathcal{F}$ for each $t \in \mathcal{T}$?*

**Exercise 4.5.** *Suppose that a gambler starts with $Z_0$ pounds, and for every $t = 1, 2, 3, \cdots$ can make a bet of $C_t$ pounds for any $0 \leq C_t \leq Z_{t-1}$, where $Z_{t-1}$ pounds is their total fortune at time $t - 1$. At each $t \geq 1$, they will win back what they bet plus the same again with probability $1/2$, or lose what they bet with probability $1/2$, so that $Z_t = Z_{t-1} + C_t$ or $Z_t = Z_{t-1} - C_t$ with equal probability. Let $(\mathcal{F}_t)_{t\geq 0}$ be the natural filtration generated by $(Z_t)_{t\geq 0}$.*

Suppose that the betting strategy is to always bet everything available. In this case, is $C_t$ a predictable process for the filtration $(\mathcal{F}_t)_{t\geq 0}$? What if the gambler decides to flip a coin, and then bet nothing if it lands tails or everything if it lands heads?

**Exercise 4.6.** *Suppose that for some* $\sup_{t\in\mathcal{T}} \mathbb{E}(|X_t|^2) < \infty$*. Show that* $(X_t)_{t\in\mathcal{T}}$ *is uniformly integrable. (Hint: you may want to use the Cauchy–Schwarz inequality).*

**Exercise 4.7.** *($\star$ Challenge $\star$) Suppose that* $(X_t)_{t=0,1,2}$ *is uniformly integrable and converges to* $X_\infty$ *in probability as* $t \to \infty$*. Show that* $X_t \to X_\infty$ *in* $L^1$*. Let* $X_1, X_2, \ldots$ *be i.i.d.* $\mathcal{N}(0,1)$ *Gaussian random variables. Set* $S_0 = 0$ *and let* $S_t = \sum_{i=1}^{t} X_i$ *for* $t \geq 1$*. Show that* $(S_t)_{t\in\mathcal{T}}$ *is a discrete time Gaussian process. Compute its mean and covariance functions.*

**Solutions to exercises from §4**

1. The discrete time SSRW has $\mathbb{E}(S_n) = \mathbb{E}(\sum_{i=1}^n X_i) = \sum_{i=1}^n \mathbb{E}(X_i) = 0$ so $\mu_S \equiv 0$. The same holds by linearity of expectation for the continuous time symmetric random walk. The expectation function of the process $(N_t)_{t\geq 0}$ is more interesting. Note that $\mathbb{P}(N_t \geq n) = \mathbb{P}(T_1 + \cdots + T_n \leq t)$, where $T_1 + \cdots + T_n$ is a sum of independent Exp(1) random variables. The distribution of $T_1 + \cdots + T_n$ is called a Gamma distribution with parameters $n$ and 1. In particular, $\mathbb{P}(T_1 + \cdots + T_n \leq t) = e^{-t}\sum_{k=t}^{\infty} t^k/k!$, which is the probability that a Poisson random variable with parameter $t$ is greater than or equal to $t$. We therefore have $N_t \sim \text{Poi}(t)$, and so $\mathbb{E}(N_t) = \sum_{t\geq 0} e^{-t}\sum_{k=t}^{\infty} t^k/k! = t$. This means that the expectation function of $N$ is given by $\mu_N(t) = t$ for $t > 0$.

   Now for the covariance functions. In the discrete time SSRW case, if $m < n$,

   $$\mathbb{E}(S_n S_m) = \mathbb{E}(\sum_{i=1}^{n} X_i \sum_{j=1}^{m} X_j) = \mathbb{E}(\sum_{i=1}^{m} X_i^2) + \mathbb{E}(\sum_{i\neq j} X_i X_j) = \sum_{i=1}^{m} \mathbb{E}(X_i^2) + 0 = m$$

   (since $\mathbb{E}(X_i X_j) = \mathbb{E}(X_i)\mathbb{E}(X_j) = 0$ for $i \neq j$ and $\mathbb{E}(X_i^2) = 1$ for each $i$. Since $\mathbb{E}(S_n)\mathbb{E}(S_m) = 0$, this implies that $\text{cov}(S_n, S_m) = m$. So,

   $$c_S(n, m) = \min\{n, m\} \text{ for } n, m \in \mathbb{N}.$$

   For the continuous time symmetric random walk we have $S_t = \sum_{i=1}^{\infty} X_i \mathbf{1}_{N_t \geq i}$ and therefore $\mathbb{E}(S_t S_s) = \mathbb{E}(\sum_{i,j=1}^{\infty} X_i X_j \mathbf{1}_{N_t\geq i}\mathbf{1}_{N_s\geq j})$ where $\mathbb{E}(X_i^2 \mathbf{1}_{N_t\geq i}\mathbf{1}_{N_s\geq i}) = \mathbb{P}(N_{\min\{s,t\}} \geq i)$ by independence of $(X_i)_i$ and $(N_t)_t$; similarly $\mathbb{E}(X_i X_j \mathbf{1}_{N_t\geq i}\mathbf{1}_{N_s\geq j}) = 0$ for $i \neq j$. Thus

   $$\mathbb{E}(S_t S_s) = \sum_{i=1}^{\infty} \mathbb{P}(N_{\min\{s,t\}} \geq i) = \mathbb{E}(N_{\min\{s,t\}}) = \min\{s, t\} \text{ for } s, t \geq 0$$

   and the covariance function of the continuous time SSRW is also given by $c_S(s, t) = \min\{s, t\}$.

2. For the Poisson process $(N_t)_{t\geq 0}$, we calculated the expectation function in the previous exercises. It also follows from the memorylessness property of the exponential distribution that if $t \geq s$: $N_s$ and $(N_t - N_s)$ are independent, and $N_t - N_s$ has the same distribution $N_{t-s}$. More generally, the memorylessness property implies that $N$ has stationary and independent increments.

   This also helps us to calculate the covariance function. Namely, it means that $\mathbb{E}(N_t N_s) = \mathbb{E}(N_s^2 + (N_t - N_s)N_s) = \mathbb{E}(N_s^2) + \mathbb{E}(N_s)\mathbb{E}(N_t - N_s)$ where $N_s \sim \text{Poi}(s)$ has $\mathbb{E}(N_s) = s$ and $\mathbb{E}(N_s^2) = s + s^2$, while $\mathbb{E}(N_t - N_s) = \mathbb{E}(N_{t-s}) = t - s$. Hence

   $$\text{cov}(N_t, N_s) = \mathbb{E}(N_t N_s) - \mathbb{E}(N_t)\mathbb{E}(N_s) = s + s^2 - s(t - s) - st = s \Rightarrow c_N(s, t) = \min s, t \text{ for } s, t \geq 0.$$

3. Strict stationarity implies that $\mathbb{E}(X_t) = \mathbb{E}(X_{t+h})$ for any $t, t + h \in \mathcal{T}$. This implies that the expectation function is constant. Suppose $s, t$ and $s', t'$ are all in $\mathcal{T}$ and $|s - t| = |s' - t'|$. Then without loss of generality we may assume that $s \leq t$, $s' = s + h$ and $t' = t + h$ for $h > 0$. Applying the strict stationarity assumption gives that $c(s, t) = c(s', t')$.

4. We know that each $\mathcal{F}_t$ is a sub $\sigma$-field of $\mathcal{F}$ for every $t$. And it is obviously increasing, so does define a filtration. But a random variable is measurable with respect to $\{\emptyset, \Omega\}$ iff it is constant, i.e. equal to some fixed value with probability one (as deduced in an earlier exercise). Thus, any process that is measurable with respect to this filtration must be a deterministic function.

   Similarly, $\mathcal{F}_t \equiv \mathcal{F}$ defines a filtration. But now, all random variables are measurable with respect to $\mathcal{F}_t$ for any $t$: thus any stochastic process is measurable with respect to this filtration.

5. In the first case, $C_t = Z_{t-1}$ so is $\mathcal{F}_{t-1}$ measurable for every $t$. This means that the process $C = (C_t)_{t=0,1,2...}$ is predictable. In the second case $\mathbb{C}_t = Z_{t-1}\mathbf{1}_A$, where $A$ is the event that the coin flipped at time $t$ is heads, and in particular is independent of $\mathcal{F}_{t-1}$. This means that $C_t$ is not $\mathcal{F}_{t-1}$ measurable (for example the event $C_t > 0$ is not in $\mathcal{F}_{t-1}$) and so $C$ is not predictable.

6. For any $K$ and $t \in \mathcal{T}$ we have $\mathbb{E}(|X_t|\mathbf{1}_{|X_t|\geq K}) \leq \mathbb{E}(|X_t|^2)^{1/2}\mathbb{P}(|X_t| \geq K)^{1/2}$ by the Cauchy–Schwarz inequality, while $\mathbb{P}(|X_t| \geq K) \leq \mathbb{E}(|X_t|^2)/K^2$ by Markov's inequality. Hence

$$\sup_{t\in\mathcal{T}} \mathbb{E}(|X_t|\mathbf{1}_{|X_t|\geq K}) \leq \sup_{t\in\mathcal{T}} \frac{\mathbb{E}(|X_t|^2)}{K}$$

which by the assumption that $\sup_{t\in\mathcal{T}} \mathbb{E}(|X_t^2|) < \infty$ does tend to 0 as $K \to \infty$. Thus $(X_t)_{t\in\mathcal{T}}$ is uniformly integrable.

# 5 Martingales

Complimentary reading: [1, §1.5].

## 5.1 Definitions

Another class of stochastic processes that are extremely important in financial mathematics (and in fact generally in probability theory) are **martingales**. Informally, these have the special property that given the values of the process up to time $T$, the value at any later time has conditional expectation equal to the value observed at time $T$. For instance, one might consider the wealth of a gambler over time if they are playing a completely fair game.

**Definition 5.1** (Continuous time martingale)**.** *Suppose that $T \in [0, \infty]$. A stochastic process $\mathfrak{X} = (X_t; t \in [0, T])$ on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in[0,T]}, \mathbb{P})$ is called a (continuous time)* ***martingale*** *if:*

*(1) it is adapted;*

*(2) $X_t$ is integrable for every $t \in [0, T]$;*

*(3) $\mathbb{E}(X_t|\mathcal{F}_s) = X_s$ a.s. for all $0 \leq s < t \leq T$.*

**Remark 5.2.** If (1) and (2) hold and $=$ is replaced with $\leq$ (respectively $\geq$) in condition (3) above, the process is called a **supermartingale** (respectively a **submartingale**).

**Definition 5.3** (Discrete time martingale)**.** *A stochastic process $\mathfrak{X} = (X_t; t = 0, 1, 2, \dots)$ on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in[0,T]}, \mathbb{P})$ is called a (discrete time)* ***martingale*** *if:*

*(1) it is adapted;*

*(2) $X_t$ is integrable for every $t \in [0, T]$;*

*(3) $\mathbb{E}(X_t|\mathcal{F}_{t-1}) = X_{t-1}$ a.s. for all $t = 0, 1, 2, \dots$.*

**Remark 5.4. Submartingales** and **supermartingales** are also defined similarly in this case.

**Remark 5.5.** An important property of martingales is that they have **constant expectation**. That is, for any $s, t \in \mathcal{T}$

$$\mathbb{E}(X_t) = \mathbb{E}(\mathbb{E}(X_t|\mathcal{F}_s)) = \mathbb{E}(X_s) \tag{93}$$

(This follows from the basic property of conditional expectation, that if $X$ is a random variable and $\mathcal{G}$ is a $\sigma$-algebra, then $\mathbb{E}(\mathbb{E}(X|\mathcal{G})) = \mathbb{E}(X)$.)

### Examples

**Example 5.6** (Random walk)**.** *A classical example of a discrete (resp. continuous) time martingale (with respect to its own natural filtration) is the discrete (resp. continuous) time random walk from the previous chapter.*

*To check this, let us consider the discrete time version $(S_t)_{t=0,1,\dots}$. If the filtration is set to be $\mathcal{F}_t = \sigma(S_0, \dots, S_t)$ for each $t \geq 0$ then $S$ is clearly adapted to this filtration. Moreover, since the walker can only take steps of size $\pm 1$, it holds that $|S_t| \leq t$ and so $S_t$ is integrable for every $t$. So it only remains to check condition (3). This follows since $S_t = \sum_{i=1}^t X_i$, which means that*

$$\mathbb{E}(S_t|\mathcal{F}_{t-1}) = \mathbb{E}(S_{t-1} + X_t|\mathcal{F}_{t-1}) = \mathbb{E}(S_{t-1}|\mathcal{F}_{t-1}) + \mathbb{E}(X_t|\mathcal{F}_{t-1}) = S_{t-1} + \mathbb{E}(X_t) = S_{t-1} \tag{94}$$

*with probability one. The second inequality here follows by linearity of conditional expectation, and the third since $S_{t-1}$ is measurable with respect to $\mathcal{F}_{t-1}$, while $X_t$ is independent of it. The final equality holds since $\mathbb{E}(X_t) = (-1) \times \mathbb{P}(X_t = -1) + 1 \times P(X_t = 1) = -1/2 + 1/2 = 0$.*

**Example 5.7** (Closed martingale)**.** *Suppose that* $(\Omega, \mathcal{F}, (\mathcal{F})_{t\in\mathcal{T}}, \mathbb{P})$ *is a filtered probability space, with* $\mathcal{T} = \{0, 1, 2, \ldots\}$ *or* $\mathcal{T} = [0, T]$ *for* $T \in [0, \infty]$*. Suppose that* $X$ *is an integrable random variable on* $(\Omega, \mathcal{F}, \mathbb{P})$*. Then*

$$(X_t)_{t\in\mathcal{T}} := (\mathbb{E}(X|\mathcal{F}_t))_{t\in\mathcal{T}} \tag{95}$$

*is a martingale in discrete/continuous time.*

**Theorem 5.8** (Doob–Meyer decomposition theorem)**.** *Suppose that* $\mathcal{T} = \{0, 1, \ldots\}$ *and that* $\mathfrak{X} = (X_t)_{t\in\mathcal{T}}$ *is a stochastic process on* $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in\mathcal{T}}, \mathbb{P})$*. Then* $\mathfrak{X}$ *is a submartingale if and only if it is the sum of a martingale and an increasing predictable process.*

*Proof.* Suppose first that $\mathfrak{X} = (X_t)_{t=0,1,\ldots}$ is a submartingale. Define

$$A_0 = 0\ ;\ A_t = \mathbb{E}(X_t - X_{t-1}|\mathcal{F}_{t-1}) \text{ for } t \geq 1\ ;\ M_t = X_t - A_t \text{ for } t \geq 0. \tag{96}$$

Then it is clear by linearity of conditional expectation that $M_t$ is a martingale. So it remains to show that $A$ is an increasing predictable process. The fact that $A$ is predictable follows from the definition, since $A_t$ is a conditional expectation with respect to $\mathcal{F}_{t-1}$, so must therefore be $\mathcal{F}_{t-1}$-measurable. The fact that $A$ is increasing follows from the assumption that $\mathfrak{X}$ is a submartingale.

Conversely, if $X = M + A$ is the sum of a martingale and a predictable, increasing process then it is easy to check that

$$\mathbb{E}(X_t|\mathcal{F}_{t-1}) = \mathbb{E}(M_t|\mathcal{F}_{t-1}) + \mathbb{E}(A_t|\mathcal{F}_{t-1}) = M_{t-1} + A_t \geq M_{t-1} + A_{t-1} = X_{t-1}, \tag{97}$$

and so $X$ is a submartingale. □

### Exercises from §5.1

**Exercise 5.1.** *Verify that the process in Example 5.7 is a martingale, with respect to the filtration* $(\mathcal{F}_t)_{t\in\mathcal{T}}$*, using the tower property of conditional expectation.*

**Exercise 5.2.** *Suppose that* $(X_t)_{t\in\mathcal{T}}$ *is a martingale (in discrete or continuous time) and that* $\phi : \mathbb{R} \to \mathbb{R}$ *is a convex function such that* $\phi(X_t)$ *is integrable for every* $t$*. Show, using conditional Jensen's inequality, that* $(\phi(X_t))_{t\in\mathcal{T}}$ *is a submartingale.*

## 5.2 Applications

**Optional stopping**

The definition of a martingale $(X_t)_{t\in\mathcal{T}}$ means that, for any $t = 0, 1, \ldots$ (discrete case) or in $[0, T]$ (continuous case), the expectation $\mathbb{E}(X_t)$ is constant. In fact this holds for a more general class of times $t$, that are allowed to be random. For a time $\tau$ to be in this class, it essentially suffices that for every $t \in \mathcal{T}$, the filtration (that $X$ is a martingale with respect to) up to time $t$ "knows" whether or not $\tau$ has already happened.

**Definition 5.9** (Stopping time)**.** *Suppose that* $\mathcal{T} = [0, T]$ *for* $T \in [0, \infty]$ *or* $\mathcal{T} = \{0, 1, \ldots\}$*. Let* $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in\mathcal{T}}, \mathbb{P})$ *be a filtered probability space. Then a random variable* $\tau : \Omega \to \mathcal{T}$ *is a* ***stopping time*** *(for this filtered probability space) if the event*

$$\{\tau \leq t\} \in \mathcal{F}_t \text{ for all } t \in \mathcal{T}. \tag{98}$$

**Remark 5.10.** In the discrete case this is equivalent to the property that $\{\tau = t\} \in \mathcal{F}_t$ for every $t = 0, 1, \ldots$.

**Example 5.11.** *Consider the discrete time random walk from Example 5.6. Let $\tau = \min\{t : S_t = 10\}$ be the first time that the walk hits 10: this is called a (first)* ***hitting time****. This is an example of a stopping time (for the natural filtration of the random walk) because for any $t = 0, 1, \ldots$, the event $\{\tau = t\}$ depends only on $S_0, S_1, \ldots, S_t$. Since these are all measurable with respect to $\mathcal{F}_t$ it follows that $\{\tau = t\} \in \mathcal{F}_t$.*

As explained above, if $(X_t)_{t\in\mathcal{T}}$ and $\tau$ are a martingale and a stopping time respectively, with respect to the same filtration, then the equality $\mathbb{E}(X_\tau) = \mathbb{E}(X_0)$ will hold. In fact, it is possible to say more:

**Theorem 5.12** (Optional stopping theorem; discrete case)**.** *Suppose that $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in 0,1,\ldots}, \mathbb{P})$ is a filtered probability space, that $(X_t)_{t\in 0,1,\ldots}$ is a martingale on it, and that $\tau$ is a stopping time. Then*

- *the process $(X^\tau_t)_{t\in 0,1,\ldots} := (X_{\tau\wedge t})_{t\in 0,1,\ldots}$ is a martingale, in particular $\mathbb{E}(X_{\tau\wedge t}) = \mathbb{E}(X_0)$ for all $t$;*
- *if $\tau$ is bounded then $\mathbb{E}(X_\tau) = \mathbb{E}(X_0)$;*
- *if $\tau < \infty$ a.s. and $|X_t| \le Y$ a.s. for all $t \in 0, 1, \ldots$, where $Y$ is an integrable random variable, then $\mathbb{E}(X_\tau) = \mathbb{E}(X_0)$.*

*Proof.* First, $(X^\tau_t)_{t\in 0,1,\ldots}$ is adapted, because $X^\tau_t = X_{t\wedge\tau}$, so for any $A \in \mathfrak{B}(\mathbb{R})$, $\{X^\tau_t \in A\} := (\cup_{m=0}^t \{X_m \in A\} \cap \{\tau = m\}) \bigcup (\{X_t \in A\} \cap \{\tau > t\}) \in \mathcal{F}_t$. It is also integrable, because for any $t$ $\mathbb{E}(|X^\tau_t|) \le \mathbb{E}(\max_{0\le m\le t} |X_m|) \le \sum_{m=0}^t \mathbb{E}(|X_m|) < \infty$. Finally, the martingale property holds, since for $t \ge 1$:

$$\mathbb{E}(X^\tau_t|\mathcal{F}_{t-1}) = \mathbb{E}(\sum_{m=0}^{t-1} X_m \mathbf{1}_{\tau=m} + X_t \mathbf{1}_{\tau\ge t}|\mathcal{F}_{t-1}) = \sum_{m=0}^{t-1} X_m \mathbf{1}_{\tau=m} + \mathbf{1}_{\tau\ge t}\mathbb{E}(X_t|\mathcal{F}_{t-1}) \tag{99}$$

where the last equality holds by the basic properties of conditional expectation and "taking out what is known" with $\{\tau \ge t\} = \{\tau \le t-1\}^c \in \mathcal{F}_{t-1}$. Since $X$ is assumed to be a martingale this is equal to

$$\sum_{m=0}^{t-2} X_m \mathbf{1}_{\tau=m} + X_{t-1}\mathbf{1}_{\tau\ge t-1} = X^\tau_{t-1}, \tag{100}$$

so $X^\tau$ is indeed a martingale.

For the second point, if $\tau$ is bounded, then there exists some $n$ such that $\mathbb{P}(\tau \le n) = 1$. So for this $n$,

$$\mathbb{E}(X_\tau) = \mathbb{E}(X_{\tau\wedge n}) = \mathbb{E}(X^\tau_n) = \mathbb{E}(X^\tau_0) = \mathbb{E}(X_0), \tag{101}$$

by Remark 5.5 since $X^\tau$ is a martingale.

Finally, if $\tau < \infty$ and $|X_t| \le Y$ for all $t$ a.s. where $Y$ is integrable, then by the dominated convergence theorem:

$$\mathbb{E}(X_\tau) = \mathbb{E}(\lim_{t\to\infty} X_{\tau\wedge t}) = \lim_{t\to\infty} \mathbb{E}(X_{\tau\wedge t}) = \mathbb{E}(X^\tau_t) = \mathbb{E}(X^\tau_0) = \mathbb{E}(X_0). \tag{102}$$

□

**Example 5.13.** *Again consider the discrete time simple symmetric random walk $(S_t)_{t=0,1,\ldots}$ from Example 5.6. This is a martingale with respect to $(\mathcal{F}_t)_{t=0,\ldots} = (\sigma(X_1, \ldots, X_t))_{t=0,\ldots} = (\sigma(X_1, \ldots, X_t))_{t=0,\ldots}$.*

*Let $a, b > 0$ be given, and let us try to compute the probability that the random walk hits $-a$ before $b$. That is, if $T_{-a}$ and $T_b$ are the first hitting times of $-a$ and $b$ respectively by the random walk, then the probability $\mathbb{P}(T_{-a} < T_b)$. To calculate this, since $\mathbb{E}(\tau) := \mathbb{E}(T_{-a} \wedge T_b) < \infty$ and the increments of $S$ are bounded, the above exercise means that the optional stopping theorem can be applied. Consequently, it holds that*

$$\mathbb{E}(S_\tau) = E(S_0) = 0. \tag{103}$$

*On the other hand,*

$$\mathbb{E}(S_\tau) = \mathbb{E}(S_\tau \mathbf{1}_{\tau=T_{-a}} + S_\tau \mathbf{1}_{\tau=T_b}) = (-a)\mathbb{P}(T_{-a} < T_b) + b\mathbb{P}(T_b < T_{-a}). \tag{104}$$

*Putting these together and writing $x = \mathbb{P}(T_{-a} < T_b) = 1 - \mathbb{P}(T_b < T_{-a})$, it follows that*

$$0 = -a.x + b(1-x) \quad \Rightarrow \quad \mathbb{P}(T_{-a} < T_b) = x = \frac{b}{a+b}. \tag{105}$$

**Theorem 5.14.** *Suppose that $\mathcal{T} = [0, T]$ for some $T \in [0, \infty]$, that $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in\mathcal{T}}, \mathbb{P})$ is a probability space, that $\mathfrak{X} = (X_t)_{t\in\mathcal{T}}$ is a martingale on it and that $\tau$ is a stopping time. Suppose that $\mathfrak{X}$ is a.s. cadlag and uniformly integrable. Then*

$$\mathbb{E}(X_\tau) = \mathbb{E}(X_0). \tag{106}$$

*Proof.* Omitted. A proof can be found, for example, in the lecture notes of Prof. J. Norris: available online at http://www.statslab.cam.ac.uk/~james/Lectures/ap.pdf (Theorem 4.3.8). □

## Martingale difference and transform

**Definition 5.15** (Martingale difference)**.** *Let $\mathfrak{X} = (X_t)_{t=0,1,\dots}$ be a discrete time martingale on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t=0,1,\dots}, \mathbb{P})$. Then the process*

$$Y_t := X_t - X_{t-1} \qquad t = 1, 2, 3, \dots \tag{107}$$

*is called a* ***martingale difference sequence*** *on $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t=0,1,\dots}, \mathbb{P})$.*

**Definition 5.16** (Martingale transform)**.** *Suppose that $Y$ is a martingale difference sequence and that $(C_t)_{t=1,2,\dots}$ is a predictable process on $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t=0,1,\dots}, \mathbb{P})$. Then the process $(Z_t)_{t=0,1,\dots}$ defined by*

$$Z_0 = 0; \quad Z_t = \sum_{i=1}^{t} C_i Y_i \tag{108}$$

*is called the* ***martingale transform*** *of $Y$ by $C$, and denoted*

$$Z = C \bullet Y. \tag{109}$$

**Lemma 5.17.** *If $C, Y, Z$ are as above and $\mathbb{E}(C_t^2), \mathbb{E}(Y_t^2) < \infty$ for all $t$, then $(Z_t)_{t=0,1,\dots}$ is a martingale with respect to $(\mathcal{F}_t)_{t=0,1,\dots}$.*

*Proof.* First, since $C_s, Y_s$ are $\mathcal{F}_t$ measurable for $s \le t$, the sum $Z_t = \sum_{i=1}^t C_i Y_i$ is $\mathcal{F}_t$-measurable for each $t$. So $Z$ is adapted to the filtration in question. The fact that is integrable can be seen by induction, since $Z_0$ is clearly integrable, and if $\mathbb{E}(|Z_s|) < \infty$ for $s = 0, 1, \dots, t$ then $\mathbb{E}(|Z_{t+1}|) \le \mathbb{E}(|Z_t| + |C_t Y_t|)$ where $\mathbb{E}(|C_t Y_t|) \le \mathbb{E}(|C_t|^2)\mathbb{E}(|Y_t|^2) < \infty$ by assumption and the Cauchy–Schwarz inequality. Finally, for any $t \ge 1$

$$\mathbb{E}(Z_t|\mathcal{F}_{t-1}) = \mathbb{E}(Z_{t-1} + C_t Y_t|\mathcal{F}_{t-1}) = Z_{t-1} + C_t \mathbb{E}(Y_t|\mathcal{F}_{t-1}) = Z_{t-1}, \tag{110}$$

which shows the martingale property. The first equality above holds since $Z_{t-1}, C_t$ are $\mathcal{F}_{t-1}$-measurable (using the "taking out what is known" lemma), and the second holds by Exercise 5.7. □

**Example 5.18.** *Take the set up of the gambler exercise from §4.2, and suppose that $(C_t)_{t=1,2,\dots}$ is predictable betting strategy, meaning that $C_t$ is $\mathcal{F}_{t-1} = \sigma(Z_0, Z_1, \dots, Z_{t-1})$-measurable for every $t$. Then $(Z_t - Z_0)$ is the martingale transform of $(Y_t)_{t\ge 0}$ by $(C_t)_{t\ge 0}$ where for each $t$, $Y_t$ is $+1$ if the bet is won, and $-1$ if it is lost. Note that $Y_t$ is the martingale difference sequence associated with the martingale $X_t = \sum_{i=1}^t Y_i$.*

### Exercises from §5.2

**Exercise 5.3.** *Show that* $\tau = \max\{1 \le t \le 100 : S_t = 10\}$ *(with* $\tau = 0$ *if* $S_t \ne 10\ \forall 1 \le t \le 100$*) is* ***not*** *a stopping time for the random walk in Example 5.11 (for the same filtration considered there).*

**Exercise 5.4.** *Take the same set-up as in the Theorem 5.12. Suppose that* $\tau$ *is integrable and* $(X_t)_{t\in 0,1,\ldots}$ *has bounded increments: that is, for some* $M < \infty$, $|X_{t+1} - X_t| \le M$ *for all* $t$ *a.s. Under this condition show that* $\mathbb{E}(X_\tau) = \mathbb{E}(X_0)$.

**Exercise 5.5.** *(⋆ Challenge ⋆) Again take the same set-up as in Theorem 5.12, and suppose that* $\mathfrak{X}$ *is uniformly integrable,* $\mathbb{P}(\tau < \infty) = 1$*. Show that* $\mathbb{E}(X_\tau) = \mathbb{E}(X_0)$.

**Exercise 5.6.** *(**Wald's identity**) Suppose that* $X_1, X_2, \ldots$ *is an i.i.d. sequence of random variables with* $X_t$ *integrable and* $\mathbb{E}(X_t) = \mu < \infty$ *for each* $t$*. Write* $S_0 = 0$, $S_t = \sum_{m=1}^{t} X_m$ *for* $t > 0$*. Show that for any* $t \ge 1$, $\mathbb{E}(S_t) = t\mu$*. Show more generally that if* $\tau$ *is a stopping time with* $\mathbb{E}(\tau) < \infty$*, then*

$$\mathbb{E}(S_\tau) = \mu\mathbb{E}(\tau). \tag{111}$$

Hint: first suppose that the $(X_i)_{i\ge 1}$ are positive random variables.

**Exercise 5.7.** *Show that if* $Y$ *is as in Definition 5.15, then* $\mathbb{E}(Y_{t+1}|\mathcal{F}_t) = 0$ *for all* $t = 0, 1, \ldots$.

**Exercise 5.8.** *Take the set up of Example 5.18. Using Lemma 5.17, deduce how much money should the gambler expect to make.*

**Solutions to exercises from §5**

1. Since $X$ is integrable, $\mathbb{E}(X|\mathcal{F}_t)$ is well-defined and integrable for every $t$. It is also $\mathcal{F}_t$-measurable, so $X$ is adapted. Finally, we have for $s \le t$: $\mathbb{E}(X_t|\mathcal{F}_s) = \mathbb{E}(\mathbb{E}(X|\mathcal{F}_t)|\mathcal{F}_s) = \mathbb{E}(X|\mathcal{F}_s) = X_s$, where the middle equality follows by the tower law. Therefore $(\mathbb{E}(X|\mathcal{F}_t))_{t\ge 0}$ is a martingale.

2. $\phi(X)$ is integrable by assumption, and adapted since $X_t$ being measurable w.r.t. $\mathcal{F}_t$ implies that $\phi(X_t)$ is measurable for $\mathcal{F}_t$. For $s \le t$ we also have $\mathbb{E}(\phi(X_t)|\mathcal{F}_s) \le \phi(\mathbb{E}(X_t|\mathcal{F}_s)) = \phi(X_s)$ since $X$ is a martingale. This implies that $\phi(X)$ is a martingale.

3. For any $1 \le t < 100$ we have $\{\tau \le t\} = \{S_s \ne 10 : t+1 \le s \le 100\}$, which is not in $\mathcal{F}_t$, since $\{X_s : s \ge t+1\}$ is independent of $\mathcal{F}_t$ by definition.

4. We have $\mathbb{E}(X_{\tau\wedge t}) = \mathbb{E}(X_0)$ for any $t < \infty$, and also that $X_{\tau\wedge t} \to X_\tau$ as $t \to \infty$. We would like to use the dominated convergence theorem to conclude that
$$\mathbb{E}(X_{\tau\wedge t}) \to \mathbb{E}(X_\tau)$$
which provides the result, since the left-hand side is constant and equal to $\mathbb{E}(X_0)$. To apply the dominated convergence theorem, we note that by the triangle inequality
$$|X_{t\wedge\tau}| \le M(t\wedge\tau) \le M\tau$$
for every $t$, where $M\tau$ is an integrable random variable by assumption. So we may use this as the dominating random variable.

5. ($\star$ Challenge $\star$)

6. The fact that $\mathbb{E}(S_t) = t\mu$ just follows by linearity of expectation. Furthermore, we have that $S_t - \mu t$ is actually a martingale with respect to the filtration defined by $\mathcal{F}_t = \sigma(X_1, \dots, X_t)$ for every $t$. Indeed, it is adapted and integrable by definition (and the fact that the $X_i$ are integrable) and for every $t$ we have
$$\mathbb{E}(S_t - \mu t|\mathcal{F}_{t-1}) = \mathbb{E}(S_{t-1} - \mu(t-1) + X_t - \mu|\mathcal{F}_{t-1}) = S_{t-1} - \mu(t-1) + \mathbb{E}(X_t - \mu|\mathcal{F}_{t-1}) = S_{t-1} - \mu(t-1).$$
We can therefore apply the optional stopping theorem so see that for any $t < \infty$,
$$\mathbb{E}(S_{t\wedge\tau} - \mu(t\wedge\tau)) = \mathbb{E}(S_0) = 0 \quad\Rightarrow\quad \mathbb{E}(S_{t\wedge\tau}) = \mu\mathbb{E}(\tau\wedge t)\ \forall t.$$
Now let us write $X_i = X_i^+ - X_i^-$ for every $i$, with $X_i^\pm$ positive, and accordingly write $\mu = \mu^+ - \mu^- = \mathbb{E}(X_1^+) - \mathbb{E}(X_1^-)$ and $S_t^\pm = \sum_{i=1}^t X_i^\pm$, so that $\mu = \mu^+ - \mu^-$ and $S_t = S_t^+ - S_t^-$. Then the above argument gives that
$$\mathbb{E}(S_{t\wedge\tau}^+) \to \mu^+\tau \text{ as } t \to \infty$$
where $\lim_{t\to\infty} \mathbb{E}(S_{t\wedge\tau}^+) \to S_\tau^+$ by the monotone convergence theorem. Thus, $\mathbb{E}(S_\tau^+) = \mu^+\mathbb{E}(\tau)$. Similarly $\mathbb{E}(S_\tau^-) = \mu^-\mathbb{E}(\tau)$, and hence
$$\mathbb{E}(S_\tau) = \mu^+\mathbb{E}(\tau) - \mu^-\mathbb{E}(\tau) = \mu\mathbb{E}(\tau)$$
as required.

7. $\mathbb{E}(Y_{t+1}|\mathcal{F}_t) = \mathbb{E}(X_{t+1} - X_t|\mathcal{F}_t) = 0 = \mathbb{E}(X_{t+1}|\mathcal{F}_t) - \mathbb{E}(X_t|\mathcal{F}_t)$. But $X_t$ is $\mathcal{F}_t$-measurable (since $X$ is adapted), and therefore $\mathbb{E}(X_t|\mathcal{F}_t) = X_t$. On the other hand, $\mathbb{E}(X_{t+1}|\mathcal{F}_t) = X_t$, since $X$ is assumed to be a martingale. So $\mathbb{E}(Y_{t+1}|\mathcal{F}_t) = X_t - X_t = 0$.

8. $Z_t - Z_0$ is a martingale (if the betting strategy given by $C$ is nice enough) so $\mathbb{E}(Z_\tau) = \mathbb{E}(Z_0)$ at any nice enough stopping time $\tau$. This means that the gambler should not expect to make any money.

## References

[1] Mikosch, T., *Elementary stochastic calculus—with finance in view*, Advanced Series on Statistical Science & Applied Probability, Vol. 6, World Scientific Publishing Co., Inc., River Edge, NJ, 1998. https://www-worldscientific-com.ezphost.dur.ac.uk/worldscibooks/10.1142/3856

[2] Williams, R. J., *Introduction to the mathematics of finance*, Graduate Studies in Mathematics, Vol. 72, AMS, Providence, RI, 2006. https://ebookcentral.proquest.com/lib/durham/detail.action?docID=3114675